STATISTICAL CONTRIBUTIONS TO NON-EXPERIMENTAL STUDIES

Byeongyeob Choi

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2015

Approved by:

Jason P. Fine

M. Alan Brookhart

Michael G. Hudgens

Anastasia Ivanova

Til Stürmer

ABSTRACT

Byeongyeob Choi: Statistical contributions to non-experimental studies
(Under the direction of Jason P. Fine and M. Alan Brookhart)

The objective of this research is to develop the methods of statistical inference for a causal effect of an exposure on an outcome in the presence of unmeasured confounders. Instrumental variable (IV) analysis is frequently used to estimate exposure effect for the data with unmeasured confounding. For example, in a randomized clinical trial, subjects often fail to comply with their own treatment protocols and such a non-compliance may depend on unmeasured confounders. In this case, it is challenging to obtain the true treatment effect, which can be observed when all subjects comply their assigned regime. To obtain the true treatment effect, we may conduct IV analysis with a randomization indicator as an IV. In many randomized clinical trials or observational studies, incomplete outcomes such as survival times with censoring are obtained. There is a lack of IV methods for incomplete data such as survival data. Another tool to overcome unmeasured confounding is to use negative control outcomes. Negative control outcomes should satisfy specific conditions in casual relationships with the exposure, outcomes and confounders. Several studies have used negative control outcomes to determine the presence of unmeasured confounders. Especially, the approach to use negative control outcomes has been elegantly used by epidemiologists to identify unmeasured confounding in the studies of effectiveness of influenza vaccine on the elderly. However, statistical methods using negative control outcomes to obtain the estimator for causal effect of the exposure have not been investigated well. Another goal of this research is to develop improved confidence intervals for current status data. Confidence intervals for current status data have been well studied theoretically, their practical application has been limited, in part because of poor performance in small samples and in part because of

computational difficulties. The subsampling-based method and likelihood-ratio test (LRT)-based method have been shown to have better coverage probabilities than a simple Wald-based method which may perform poorly in realistic sample sizes. However, those methods are complicated and require much more computational demands compared to Wald-based method. Therefore, we propose (1) Two-stage estimation of structural instrumental variable models with incomplete data, (2) Sensitivity analysis of regression results to unobserved confounding using a negative control outcome, (3) A new instrumental variable estimator using a negative control outcome and (4) Improved confidence intervals for current status data.

ACKNOWLEDGMENTS

I would like to thank my advisors, Jason Fine and Alan Brookhart. They provided me environments, motivations and directions so that I can pursue research for my dissertation. Also, I would like to thank my committee members, Drs. Hudgens, Ivanova and Stürmer for their invaluable comments. I would like to thank my family, who always supported me unconditionally. My special thanks go to Aro Nam Choi, who has become the most beautiful wife for me, respected me, and encouraged and supported me faithfully. Last, I thank Father God in Christ Jesus, who led me thus far and saved me from death through sin.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Researchers in epidemiology often have an interest in making a causal inference for an exposure effect on an outcome. In many cases, observation studies are employed to do such an inference. However, observation studies are easily affected by unmeasured confounders, which may result in a biased estimate of the exposure effect due to residual confounding. In economic terminology, we say the exposure variable is endogenous in this case and classical statistical methods adjusting measured confounders fail to give correct inference results for the the exposure effect.

An instrumental variable (IV) analysis is designed to overcome the unmeasured confounding problem (Brookhart et al. 2006, McClellan et al. 1994, Schneeweiss et al. 2006, Stukel et al. 2007). Although the requirements of an IV depend on a particular analytic method one chooses, we can say that a variable is an IV if it satisfies the following three conditions (Brookhart et al. 2010): (i) it has a causal effect on the exposure; (ii) it has effects on the outcome only through the exposure; (iii) it is unrelated to an unmeasured confounder. A typical example of an IV is a randomized indicator, which is usually available, to estimate a drug effect in randomized trial with non-compliance.

There has been considerable work for IV methods with complete data, where both outcome and exposure are fully observed and assumed to satisfy semiparametric linear or nonlinear structural equation models with unspecified error distributions. The most popular method is the two-stage least squares (2SLS) for linear models, which is obtained by applying least squares to the model where the endogenous variables are replaced by their predicted values (Theil 1953). Amemiya

(1974) proposed nonlinear two-stage least squares (NL2SLS), which generalizes 2SLS for nonlinear models. Newey (1990) pointed out the asymptotic efficiency of IV estimators of nonlinear model depends on the form of instruments. He proposed efficient instrumental variables estimation of nonlinear models, which employs nonparametric estimation of the optimal instruments. Abadie (2003) introduced a new class of IV estimators for linear and nonlinear treatment response models. Terza et al. (2008) introduced two-stage predictor substitution (2SPS) estimation and two-stage residual inclusion estimation (2SRI) for nonlinear models. 2SPS is the extension to nonlinear models of 2SLS. 2SRI is similar to 2SPS except that first stage residuals are included as additional regressors instead of replacing the endogenous variables by their predicted values. They showed that in a generic parametric framework, 2SRI is consistent and 2SPS is not. In case where either response or exposure is incompletely observed, however, such semiparametric methods are not applicable. An important special case of coarsened data occurs with time to event outcomes and exposures, which may be subject to both censoring and truncation.

In the case where we do not have a tool such a valid IV, thus it is hard do find a reliable way to consistently estimate the parameters, it is desirable to evaluate the sensitivity of regression results to unmeasured confounders. There have been several developed sensitivity analysis techniques (Rosenbaum and Rubin 1983, Lin et al. 1998, Brumback et al. 2004, Gustafson et al. 2010, VanderWeele et al. 2012). Among those, the methods of Lin et al. (1998) are applicable to general regression models and can be easily performed (VanderWeele 2008). Lin et al. (1998) assumed that the distribution of the unmeasured confounder conditional on the measured confounders and the exposure is normal or binomial, and they identified simple algebraic relationships between a true exposure effect in the full model and an apparent exposure effect in the reduced model which does not control for unmeasured confounders. One can make an inference on the true exposure effect by making a simple adjustment to the estimate and the confidence interval of the apparent exposure effect. Lin et al. (1998) developed their method for linear, log-linear, logistic and proportional hazard models.

Negative controls technique has been emerging as a tool to detect unmeasured confounding.

(Flanders et al. 2011, Lipsitch et al. 2010, Lumley and Sheppard 2000, Jackson et al. 2006, Smith 2008; 2012). An outcome is said to be a valid negative control outcome (N) if it is influenced by measured confounders (W) and an unmeasured confounder (U) in the association between the exposure (Z) and the main outcome (Y), but not directly influenced by the exposure (Lipsitch et al. 2010). Those conditions are sufficient to detect unmeasured confounding, but insufficient to estimate the causal effect. Tchetgen Tchetgen (2014) made a more progress to estimate the causal effect by imposing an additional assumption that the negative control outcome is independent of a treatment received conditional on the measured confounders and counterfactual outcomes. This assumption implies that the counterfactual outcomes are ideal proxies of the unmeasured confounders. Under this assumption, an additive causal effect on a continuous outcome can be estimated by regressing N onto (W, Y, Z).

Herein, we develop IV methods to estimate the causal effect of the exposure in coarsened data. The developed methods are focused on right-censored outcome data. Further, we develop a new IV estimator with a negative control outcome, which is consistent even if the IV assumption that the IV should be independent of an unmeasured confounder is violated. For the case where the IV is not available, we develop the method for sensitivity analysis with a negative control outcome under unmeasured confounding. Finally, we introduce improved confidence intervals for current status data. The remainder of this chapter provides the review of the IV methods for linear and nonlinear models with complete and incomplete data and that of the methods with a negative control outcome.

In Chapter 2, a general class of two-stage minimum distance estimators for coarsened data is proposed that separately fits the linear models for the outcome and the exposure and then estimates the true exposure effect on the outcome using a reduced form model. An optimal minimum distance estimator is identified and shown to be theoretically superior to the usual two- stage least squares estimator with fully observed data. Simulation studies demonstrate that the methods perform well with realistic sample sizes. The practical utility of the methods is illustrated in a study of the comparative effectiveness of colon cancer treatments, where the effect of chemotherapy on censored survival times is confounded by patient status. In Chapter 3, we propose a new IV estimator using a negative control outcome. The new IV estimator has been shown to be theoretically valid even if the IV independence assumption that the IV should be independent of the unmeasured confounder, is violated. Structural equation models are used to define the IV and the negative control outcome. For the case of multiple IVs, the new IV method depends on weights and the optimal weight is identified. In addition, a Wald test is proposed to test the IV independence assumption. An IV estimator combining the new IV estimator and an usual IV estimator is proposed under the IV independence assumption and shown to be superior to both individual IV estimators. In Chapter 4, we develop new methods of sensitivity analysis in a regression model with potential unmeasured confounding using a negative control outcome. As in Lin et al. (1998), we assume that the causal exposure effect is represented by a regression model. We use outcome models which follow the definition of the negative control outcome presented in Lipsitch et al. (2010) and definition 1 of Tchetgen Tchetgen (2014), but the outcomes in our models are not ideal proxies because of error terms and thus the methods of Tchetgen Tchetgen (2014) are not applicable. First, we extended the control outcome calibration approach of Tchetgen Tchetgen (2014) to conduct sensitivity analysis. Second, we extended Lin's methods to use a negative control outcome with fewer sensitivity parameters. Furthermore, we developed Lin's methods for probit model and additive hazard model and extended those to use a negative control outcome. Third, we applied Lin's methods to a conditional likelihood with a negative control outcome for binary and right-censored outcome data. The proposed methods are shown to perform well and better than Lin's method in logistic and proportional hazard regression models, and the conditional likelihood method is superior than both the marginal model method and the Lin's method when the unmeasured confounding is huge. In Chapter 5, we discuss the confidence intervals for current status data and show that by using transformations, simple Wald-based CIs can be improved with small and moderate sample sizes to have competitive performance with likelihood ratio test (LRT)-based method. Our simulations further show that a simple nonparametric bootstrap gives approximately correct CIs for the data generating mechanisms that we consider.

1.2 Instrumental variable methods

1.2.1 Complete data

First, we discuss linear models. For $i = 1, ..., n$, let y_i be the response, let z_i be the exposure which is endogenous, let $\mathbf{v}_i$ be the $p \times 1$ vector of the instrumental variables (IVs) and let $\mathbf{w}_i$ be the $q \times 1$ vector of the confounders. The linear model for the response is given by

$$y_i = \alpha_o + \alpha_{yz} z_i + \boldsymbol{\alpha}_{yw}^T \mathbf{w}_i + \varepsilon_i. \tag{1.1}$$

In this model, α_{yz} represents the average causal effect of z on y. The usual estimation methods do not give consistent estimators for the regression parameters in (1.1) because z and ε are correlated. The linear model for the exposure is given by

$$z_i = \beta_o + \boldsymbol{\beta}_{zv}^T \mathbf{v}_i + \boldsymbol{\beta}_{zw}^T \mathbf{w}_i + \eta_i.$$

In matrix notation, we can write

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha}_0 + \boldsymbol{\varepsilon},$$

$$\mathbf{z} = \mathbf{D}\boldsymbol{\beta}_0 + \boldsymbol{\eta},$$

where $\boldsymbol{\alpha}_0 = (\alpha_o, \alpha_{yz}, \boldsymbol{\alpha}_{yw})^T$, $\boldsymbol{\beta}_0 = (\beta_o, \boldsymbol{\beta}_{zv}, \boldsymbol{\beta}_{zw})^T$, and $\mathbf{y}$, $\mathbf{z}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\eta}$ are n-vectors with typical element y_i, z_i, ε_i and η_i, and $\mathbf{X}$ and $\mathbf{D}$ are $n \times (2 + q)$, $n \times (1 + p + q)$ matrices with row $(1, z_i, \mathbf{w}_i)$ and $(1, \mathbf{v}_i, \mathbf{w}_i)$.

We assume that $E(\varepsilon|\mathbf{v}) = E(\eta|\mathbf{v}) = 0$, $var(\varepsilon|\mathbf{v}) = \sigma_\varepsilon^2$, $var(\eta|\mathbf{v}) = \sigma_\eta^2$ and $E(\varepsilon\eta|\mathbf{v}) = \sigma_{\varepsilon\eta}$. We also assume that the probability limit of $\mathbf{D}^T\mathbf{D}/n$ and $\mathbf{X}^T\mathbf{X}/n$ are given by Σ_D and Σ_X respectively.

The two-stage least squares (2SLS) estimator for $\boldsymbol{\alpha}_0$ is

$$\hat{\boldsymbol{\alpha}}_{2sls} = (\mathbf{X}^T\mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{y}).$$

The 2SLS can be also obtained by applying least squares to (1.1), where z_i is replaced by its predicted value. The 2SLS is the value of $\boldsymbol{\alpha}$ to minimize

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})^T\mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}),$$

The 2SLS is consistent and the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\alpha}}_{2sls} - \boldsymbol{\alpha}_0)$ is normal with mean zero and variance $(\boldsymbol{\beta}^T\Sigma_D\boldsymbol{\beta})^{-1}\sigma_\varepsilon^2$. If we have one IV and no confounders, then the 2SLS for the parameter of the exposure effect, α_{yz}, is the ratio of the sample covariances, $\widehat{cov}(y,v)/\widehat{cov}(z,v)$.

Next, we discuss non-linear models. Let consider the following nonlinear regression model,

$$y_i = f(z_i, \mathbf{w}_i, \boldsymbol{\alpha}) + \varepsilon_i,$$

where f is a possibly nonlinear function in z, $\mathbf{w}$ and $\boldsymbol{\alpha}$. Endogeneity occurs because of the correlation of z and ε. For Amemiya's nonlinear two-stage least squares (NL2SLS) Amemiya (1974), f is assumed to have continuous first and second derivatives with respect to $\boldsymbol{\alpha}$. The NL2SLS estimator (Amemiya 1974) of $\boldsymbol{\alpha}$ is the value of $\boldsymbol{\alpha}$ that minimizes

$$\Phi(\boldsymbol{\alpha}) = (\mathbf{y} - \mathbf{f})^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{f}),$$

where $\mathbf{y}$ and $\mathbf{f}$ are n-component vectors whose ith elements are y_i and $f(z_i, \mathbf{w}_i, \boldsymbol{\alpha})$ respectively, and $\mathbf{H}$ is a $n \times K$ matrix of certain constants with rank $K \le (1 + p + q)$. If $\mathbf{H} = \mathbf{D}$, then NL2SLS reduces to 2SLS. The NL2SLS is consistent for the true value $\boldsymbol{\alpha}_0$ and under certain conditions $\sqrt{n}(\hat{\boldsymbol{\alpha}}_{nl2sls} - \boldsymbol{\alpha}_0)$ is normal with mean zero and variance

$$\sigma_\varepsilon^2 \left\{ \mathrm{plim}_{n\to\infty} \frac{1}{n} \dot{\mathbf{f}}(\boldsymbol{\alpha}_0)^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\dot{\mathbf{f}}(\boldsymbol{\alpha}_0) \right\}^{-1},$$

where $\dot{\mathbf{f}}(\boldsymbol{\alpha})$ is the first derivative of $\mathbf{f}(\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$.

Terza et al. (2008) assume the regression model of the response y is of the form

$$y_i = M(\alpha_{yz}z_i + \boldsymbol{\alpha}_{yw}^T\mathbf{w}_i + \alpha_{yu}u_i) + e_i, \tag{1.2}$$

where $M(\cdot)$ is a known nonlinear function, u is an unmeasured confounder and e is a random error tautologically defined as $e = y - M(\alpha_{yz}z_i + \boldsymbol{\alpha}_{yw}\mathbf{w}_i + \alpha_{yu}u_i)$ so that $E(e|z,\mathbf{w},u) = 0$. Endogeneity occurs because of the correlation between z and u. They assume the regression model of the z is of the form

$$z_i = r(\boldsymbol{\beta}_{zv}^T\mathbf{v}_i + \boldsymbol{\beta}_{zw}^T\mathbf{w}_i) + u_i, \tag{1.3}$$

where r is a known nonlinear function.

The 2SPS method is straightforward to implement. In the first stage, we obtain consistent estimates of $(\boldsymbol{\beta}_{zv},\boldsymbol{\beta}_{zw})$, $(\hat{\boldsymbol{\beta}}_{\boldsymbol{zv}},\hat{\boldsymbol{\beta}}_{\boldsymbol{zw}})$, by applying the nonlinear least squares (NLS) method to the model (1.3). Then, compute the predicted values of z_i, $\hat{z}_i = r(\hat{\boldsymbol{\beta}}_{zv}^T\mathbf{v}_i + \hat{\boldsymbol{\beta}}_{zw}^T\mathbf{w}_i)$, for $i = 1,...,n$. Correspondingly, define the residuals $\hat{u}_i = z_i - \hat{z}_i$. In the second stage, estimate the parameters $(\gamma_{yz},\boldsymbol{\gamma}_{yw})$ by applying NLS to the model

$$y_i = M(\gamma_{yz}\hat{z}_i + \boldsymbol{\gamma}_{yw}^T\mathbf{w}_i) + e_i^{\text{2SPS}}, \tag{1.4}$$

where e_i^{2SPS} is the regression error term. The resultant estimate, $(\hat{\gamma}_{yz},\hat{\boldsymbol{\gamma}}_{yw})$, is the estimate of $(\alpha_{yz},\boldsymbol{\alpha}_{yw})$.

The 2SRI estimation is performed by including the residual, $\hat{u}$, in the second stage model. We apply NLS to the following model,

$$y_i = M(\alpha_{yz}z_i + \boldsymbol{\alpha}_{yw}^T\mathbf{w}_i + \alpha_{yu}\hat{u}_i) + e_i^{\text{2SRI}}, \tag{1.5}$$

where e_i^{2SRI} is the regression error term. The consistency of 2SRI estimator can be proven by in the context of two-stage optimization estimator (Newey and McFadden 1994, White 1994, Wooldridge 2002). For linear models, 2SPS and 2SRI are equivalent to 2SLS.

1.2.2 Right-censored outcome data

Several IV methods for right-censored outcome data have been proposed. Robins and Tsiatis (1991) developed IV estimators to correct for non-compliance in randomized trials by estimating the parameters of a class of semiparametric failure time models, the rank preserving structural failure time models (RPSFTM), using a class of rank estimators. In a randomized clinical trial designed to study the effect of a drug on survival, subjects are assigned to treatment protocols. Unfortunately, some subjects often fail to comply with their assigned regimes. The method of Robins and Tsiatis (1991) allows one to estimate the true treatment effect, i.e., the effect that would be observed if all subjects complied with their assigned protocols, in the presence of such non-compliance.

For $i = 1, ..., n$, suppose z_i is a treatment indicator, $h_i(t) = \{z_i(u); 0 \leq u \leq t\}$ is a treatment history, v_i is a randomization group indicator and u_i is the survival time if the ith subject was never to receive treatment, i.e., $z_i(t) = 0$ for all t. Robins and Tsiatis (1991) assumed U is independent of the treatment arm to which the subject is assigned. In the absence of censoring, we observe the random variables $(T_i, H_i(T_i), V_i)$, where T_i is the observed failure time of the ith subject.

A RPSFTM relates U_i to $\{T_i, H(T_i)\}$ by assuming

$$U_i = \psi(T_i, H_i(T_i), \boldsymbol{\alpha}_0),$$

where $\boldsymbol{\alpha}_0 \in \mathbb{R}^p$ is an unknown parameter and $\psi(\cdot)$ is a known smooth function. A simple form of $\psi(T_i, H_i(T_i), \boldsymbol{\alpha}_0)$ is $\int_0^{T_i} \exp(\alpha_0 Z_i(x))dx$. The $\psi(\cdot)$ is assumed to satisfy the following conditions.

<u>Smoothness</u>: $\psi_t(t, h(t), \boldsymbol{\alpha})$, $\psi_{\alpha_j}(t, h(t), \boldsymbol{\alpha})$ and $\psi_{t,\alpha_j}(t, h(t), \boldsymbol{\alpha})$ are continuous for all $\alpha_j(j =$

$1, ..., p$) and almost all t with respect to Lesbesque measure where $\psi_t(t, h(t), \boldsymbol{\alpha}) = \partial\psi(t, h(t), \boldsymbol{\alpha})/\partial t$, $\psi_{\alpha_j}(t, h(t), \boldsymbol{\alpha}) = \partial\psi(t, h(t), \boldsymbol{\alpha})/\partial\alpha_j$ and $\psi_{t,\alpha_j}(t, h(t), \boldsymbol{\alpha}) = \partial\psi_t(t, h(t), \boldsymbol{\alpha})/\partial\alpha_j$.

Monotonicity: $\psi(t, h(t), \alpha) > \psi(u, h(u), \alpha)$ if $t > u$.

Identity: $\psi(t, h(t), 0) = t$

Independence and Indentification: There exists a unique α_0 such that

$$U(\alpha_0) \perp\!\!\!\perp V,$$

where $U(\alpha) = \psi(T, H(T), \alpha)$ and $A \perp\!\!\!\perp B$ means A and B are independent.

Define $N_i(u, \boldsymbol{\alpha}) = I(U_i(\boldsymbol{\alpha}) \le u)$ and $Y_i(u, \boldsymbol{\alpha}) = I(U_i(\boldsymbol{\alpha}) \ge u)$, where $I(A) = 1$ if statement A is true and $I(A) = 0$ otherwise. Given a known p-vector $\mathbf{g}(v, u, \boldsymbol{\alpha}) = \{g_1(v, u, \boldsymbol{\alpha}), ..., g_p(v, u, \boldsymbol{\alpha})\}^T$, $S_n(\boldsymbol{\alpha}, \mathbf{g})$ is defined to be the p-vector with components

$$S_{n,j}(\boldsymbol{\alpha}, \mathbf{g}) = \sum_{i=1}^{n} \int dN_i(u, \boldsymbol{\alpha})\{g_j(V_i, u, \boldsymbol{\alpha}) - \bar{g}_j(u, \boldsymbol{\alpha})\},\ j = 1, ..., p,$$

where

$$\bar{g}_j(u, \boldsymbol{\alpha}) = \sum_{i=1}^{n} g_j(V_i, u, \boldsymbol{\alpha}) Y_i(u, \boldsymbol{\alpha}) / \sum_{i=1}^{n} Y_i(u, \boldsymbol{\alpha}).$$

Let $\hat{\boldsymbol{\alpha}}(\mathbf{g})$ be a value of $\boldsymbol{\alpha}$ that solves $S_n(\boldsymbol{\alpha}, \mathbf{g}) = 0$. Since $S_n(\boldsymbol{\alpha}, \mathbf{g})$ is a step function in $\boldsymbol{\alpha}$, we obtain $\hat{\boldsymbol{\alpha}}(\mathbf{g})$ by minimizing $\hat{\boldsymbol{\alpha}}(\mathbf{g})^T \hat{\boldsymbol{\alpha}}(\mathbf{g})$. The consistency and asymptotic normality of $\hat{\boldsymbol{\alpha}}(\mathbf{g})$ can be proven by the asymptotic theory of rank estimators of a linear regression (Tsiatis 1990).

Robins and Tsiatis (1991) considered the restrictive type of censoring. They assumed the censoring time C is known for all subjects. The observable variables $(T_i, H(T_i), V_i, C_i)$ are assumed to be i.i.d. and $(U(\boldsymbol{\alpha}_0), C) \perp\!\!\!\perp V$. For a valid inference on $\boldsymbol{\alpha}$, a new censoring time needs to be defined and this complicates the inference.

Bijwaard (2008) extended the method of Robins and Tsiatis (1991) to generalized accelerated failure time (GAFT) models, which is based on transforming the failure times and assuming a distribution for these transformed failure times. Instrumental variable linear rank estimator (IVLR) is proposed, which exploits the fact that for the true GAFT, the IV does not influence the hazard of the transformed failure times. However, like the method of Robins and Tsiatis (1991), censoring times should be known for all subjects and a new censoring time needs to be defined.

Brannas (2000) assumed accelerated failure time (AFT) models. The estimators proposed are IV adaptations to the Powell symmetrically trimmed least squares and the Buckley-James estimators for right-censored data. That is, he applied those two estimation methods to the AFT models where the endogenous variable is replaced by its predicted values. Simulation studies showed that they perform well, however, the theoretical properties of these procedures were not investigated.

Loeys and Goetghebeur (2003) proposed a causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. Suppose that n independent subjects were randomized over experimental treatment $(V_i = 1)$ or control $(V_i = 0)$. For subject i on treatment (respectively control), all-or-nothing exposure to treatment Z_{1i} (Z_{0i}) is observed, along with its possibly right-censored survival time T_{1i} (T_{0i}). They make the following assumptions.

(A1) $(Z_{1i}, T_{1i}, Z_{0i}, T_{0i}, V_i)$ are i.i.d. random variables, implying potential outcomes for each person are unrelated to the treatment or outcome of other individuals.

(A2) The randomization assumption: $(Z_{1i}, T_{1i}, Z_{0i}, T_{0i}) \perp\!\!\!\perp V_i$.

(A3) No access to experimental therapy on the control arm. Hence, Z_{0i} equals to 0 for all subjects, and T_{0i} represents the treatment-free outcome, when randomized to control.

(A4) Following (A3), we also require that $P(T_{1i} > t|Z_{1i} = 0) = P(T_{0i} > t|Z_{1i} = 0)$. This assumption is called the “absence of indirect effect" by Pearl (2002) and “the exclusion restriction" by Angrist et al. (1996).

They use the following causal proportional hazard model,

$$\lambda(t|V_i = 1, Z_{1i} = 1) = \lambda(t|V_i = 0, Z_{1i} = 1)\exp(\alpha_0). \tag{1.6}$$

In (1.6), $\exp(\alpha_0)$ captures the causal proportional hazards effect within the treatable subpopulation. The interest here is to estimate $\exp(\alpha_0)$.

Let C_i be a censoring time for the ith subject, $\tilde{T}_i = \min(T_i, C_i)$, $\delta_i = I(T_i \leq C_i)$, $N_i(t) = I(\tilde{T}_i \leq t, \delta_i = 1)$, $Y_i(t) = I(\tilde{T}_i \geq t)$ and filtration
$mathbbF_t = \sigma\{N_i(s), Y_i(s+), V_i, Z_i, i = 1, ..., n; 0 \geq s \geq t\}$. They model intensity process as follows.

$$
\begin{aligned}
&E\{dN_i(t)|V_i(1-Z_i) = 1, Y_i(t)\} = \lambda_{00}(t)Y_i(t)dt \\
&E\{dN_i(t)|V_i = 0, Y_i(t)\} = \{(1-\pi(t))\lambda_{00}(t) + \pi(t)\lambda_{01}(t)\}Y_i(t)dt \\
&E\{dN_i(t)|V_iZ_i = 1, Y_i(t)\} = \lambda_{01}(t)e^{\alpha}Y_i(t)dt,
\end{aligned}
$$

where $\pi(t) = P(Z_{1i} = 1|\tilde{T}_i \geq t, V_i = 0)$. For known $\pi(t)$, we could then define

$$
d\hat{\Lambda}_{01}(t) = \frac{1}{\pi(t)}\left\{\sum_i \frac{(1-V_i)dN_i(t)}{\sum_j Y_j(t)(1-V_j)} - (1-\pi(t))\sum_i \frac{V_i(1-Z_i)dN_i(t)}{Y_j(t)V_j(1-Z_j)}\right\} \tag{1.7}
$$

If we substitute (1.7) into (1.6), then we obtain the following score equation,

$$
\sum_i \int dN_i(t)\left[V_iZ_i - \left\{\sum_j V_iZ_ie^{\alpha}Y_j(t)\right\}\frac{1}{\pi(t)} \times \left\{\frac{1-V_i}{\sum_j Y_j(t)(1-V_i)} - \frac{(1-\pi(t))V_i(1-Z_i)}{\sum_j Y_j(t)(1-Z_i)V_i}\right\}\right] = 0 \tag{1.8}
$$

For given $\pi(t)$, the process (1.8) has compensator 0, and the martingale central limit theorem can be used to obtain asymptotic normality of the resultant estimator. However, $\pi(t)$ is unknown by (A3). Therefore, (1.8) should be estimated. The authors estimate $S_{01}(t)$ with monotonic decreasing assumption and its jump is used for $d\hat{\Lambda}_{01}(t)$.

Loeys et al. (2005) extended the methods of Loeys and Goetghebeur (2003) to allow more general exposure level and covariates in the model. The model they considered is

$$
\lambda(t|V_i = 1, U_i = u, \mathbf{X}_i = \mathbf{x}) = \lambda(t|V_i = 0, U_i = u, \mathbf{X}_i = \mathbf{x})\exp(\alpha_0 u). \tag{1.9}
$$

where U_i is a subject i's potential exposure to the experimental treatment if he/she were randomized to treatment and $\mathbf{X}_i$ is a covariate vector for the ith subject. The challenging here is that U_i is unobserved for all subjects in the control group ($\{V_i = 0\}$). The estimation procedures relies heavily on randomization.

We rewrite the model (1.9) in terms of survival distribution,

$$S(t|V_i = 1, U_i = u, \mathbf{X}_i = \mathbf{x}) = S(t|V_i = 0, U_i = u, \mathbf{X}_i = \mathbf{x})^{\exp(\alpha_0 u)}. \tag{1.10}$$

Now, survival probability in the control group are a mixture of unobserved compliance-specific probabilities, i.e. $S(t|V_i = 0, \mathbf{X}_i = \mathbf{x})$ equals

$$\sum_u S(t|V_i = 0, U_i = u, \mathbf{X}_i = \mathbf{x}) P(U_i = u|V_i = 0, \mathbf{X}_i = \mathbf{x}), \tag{1.11}$$

when U_i is discrete. If the model (1.10) holds, the model (1.11) also equals

$$\sum_u S(t|V_i = 1, U_i = u, \mathbf{X}_i = \mathbf{x})^{\exp(-\alpha_0 u)} P(U_i = u|V_i = 1, \mathbf{X}_i = \mathbf{x}),$$

since $P(U_i = u|V_i = 0, \mathbf{X}_i = \mathbf{x}) = P(U_i = u|V_i = 1, \mathbf{X}_i = \mathbf{x})$ by definition of U_i and randomization. We define $\hat{S}_{1\to 0}(t|\mathbf{x};\alpha)$ as

$$\hat{S}(t|V_i = 1, U_i = u, \mathbf{X}_i = \mathbf{x})^{\exp(-\alpha_0 u)} \hat{P}(U_i = u|V_i = 1, \mathbf{X}_i = \mathbf{x}).$$

The unknown parameter α is estimated by the value of α that minimizes the distance between the $\hat{S}_{1\to 0}(t|\mathbf{x};\alpha)$ and the fitted treatment-free survival distribution in the control group conditional on $\mathbf{X}$. To this end, they propose a logrank test which is built as a sum of $\mathbf{x}$-specific pseudo martingales in the control group:

$$\sum_{V_i=0} \{\delta_i - \hat{\Lambda}_{1\to 0}(\tilde{T}_i|\mathbf{x}_i;\alpha)\} \tag{1.12}$$

where $\hat{\Lambda}_{1\to 0}(\tilde{T}_i|\mathbf{x}_i;\alpha) = -\log \hat{S}_{1\to 0}(\tilde{T}_i|\mathbf{x}_i;\alpha)$. The variance of (1.12) can be estimated by $2\sum_{V_i=0}\hat{\Lambda}_{1\to 0}(\tilde{T}_i|\mathbf{x}_i;\alpha)$. The point estimator for α_0 can be found as the α-value that minimizes the χ^2 value of the test statistic

$$\left(\sum_{V_i=0}\{\delta_i - \hat{\Lambda}_{1\to 0}(\tilde{T}_i|\mathbf{x}_i;\alpha)\}\right)^2 / \left(2\sum_{V_i=0}\hat{\Lambda}_{1\to 0}(\tilde{T}_i|\mathbf{x}_i;\alpha)\right).$$

The proposed methods in Chapter 2 is based on AFT model like Robins and Tsiatis (1991), Bijwaard (2008) and Brannas (2000). Unlike Brannas (2000), our methods are theoretically well justified with consistency and asymptotic normality. Furthermore, our methods do not need the condition that the consorting times are known for all subjects as in Robins and Tsiatis (1991) and Bijwaard (2008).

1.3 Negative control outcomes

1.3.1 Detection of unmeasured confounding

Lipsitch et al. (2010) provides a comprehensive review of the use of negative controls to identify confounding in observations studies. They establish the conditions under which negative control outcomes and negative control exposures can be used to detect unmeasured confounders. Here we focus on the negative control outcomes.

A negative control outcome (N) should be an outcome that are affected by the set of measured (W) and unmeasured (U) confounders of the association between the exposure (Z) and the outcome (Y). We say such N is an “U-comparable" outcome. If N is not caused by Z, then any association of N and Z observed by the same analysis method which is used to determine the association of Y and Z would indicate the bias in that of Y and Z.

As we mentioned earlier, negative controls have been elegantly used by epidemiologists to identify unmeasured confounding in studies of influenza vaccination effectiveness by Jackson et al. (2006). The negative control outcome used in that study was the mortality or pneumonia/influenza hospitalization in the period before and after influenza season. That negative control outcome

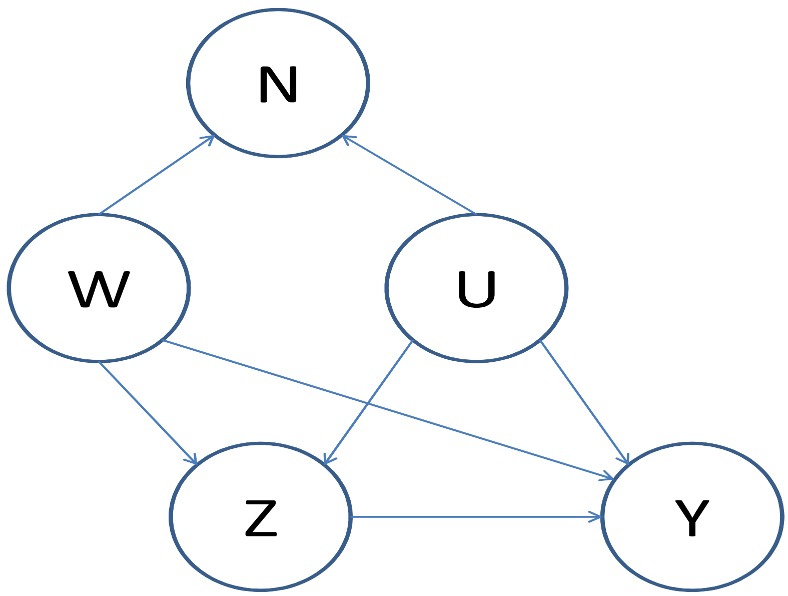

Figure 1.1: Causal diagram showing an ideal negative control outcome N for use in identifying potential unmeasured confounding. N should have the same incoming arrow as outcome Y, except that N is not caused by Z.

may satisfy “U-comparable" because protective effect of vaccination should be specific to influenza season and it may share the common confounders with mortality or pneumonia/influenza hospitalization during influenza season.

Jackson et al. (2006) also used irrelevant outcomes to influenza vaccination such as hospitalization for injury or trauma as negative control outcomes. They postulated that the effect of influenza vaccination is specific to the outcome related to influenza. They found that there was also protective effects on injury or trauma hospitalization. This indicates that the observed effectiveness of vaccination is biased due to uncontrolled confounders.

1.3.2 Control outcome calibration

Tchetgen Tchetgen (2014) developed control outcome calibration approach to estimate an exposure effect on an outcome under unmeasured confounding. His key assumption is that a negative control outcome is independent of treatment selection conditional on measured confounders and counterfactual outcomes. This assumption implies that counterfactual outcomes are ideal proxy measures for an unmeasured confounder.

Let Y_z denote a main outcome for the subject who received the treatment z. Also, let N_z denote a negative control outcome for the subject who received the treatment z. Definition 1 of Tchetgen Tchetgen (2014) says that N is a negative control outcome if $N_z = N$ $(z = 0, 1)$ for all individuals and the confounding variables for the exposure-negative control outcome association are the same as those for the exposure-main outcome association.

Let $\mathbb{Y}_Z = \{Y_z : z \in \mathbb{Z}\}$ denote the set of all counterfactuals for the main outcome under all possible values of the exposure in the set $\mathbb{Z}$. Then, Assumption 1 of Tchetgen Tchetgen (2014) says that the exposure is independent of N_z conditional on $\{W, \mathbb{Y}_Z\}$, or

$$N = N_z \perp\!\!\!\perp Z \mid \{W, \mathbb{Y}_Z\}.$$

Let $\Psi_0 = Y^1 - Y^0$ be the parameter of interest and $Y(\Psi) = Y - \Psi Z$. Under Assumption 1 of

Tchetgen Tchetgen (2014),

$$N \perp\!\!\!\perp Z \mid \{W, Y(\Psi)\},$$

if and only if $\Psi = \Psi_0$. Using a linear regression model, we can set

$$\begin{aligned} E(N \mid Z, W, Y(\Psi_0) &= E(N \mid W, Y(\Psi_0)) \\ &= \beta_1 + \beta_2^T W + \beta_3 Y(\Psi_0), \\ &= \beta_1 + \beta_2^T W + \beta_3 Y + \beta_4 Z, \end{aligned}$$

where $\beta_4 = -\beta_3 \Psi_0$, assuming that $\beta_3 \neq 0$. Thus the estimator of Tchetgen Tchetgen (2014) is give by

$$\hat{\Psi} = -\hat{\beta}_4 / \hat{\beta}_3.$$

CHAPTER 2: TWO-STAGE ESTIMATION OF STRUCTURAL INSTRUMENTAL VARIABLE MODELS WITH COARSENED DATA

2.1 Introduction

Observational studies are subject to confounding by variables which affect an exposure and an outcome. Confounding is one of major reason to yield biased estimates in observational studies. Regression adjustment or propensity score methods are usually used to overcome this problem. However, those methods require that all of the confounders are observed and this may not be the case in many cases. In economic terminology, we say the exposure variable is endogenous when the exposure is correlated with an error term by sharing unmeasured confounders. Endogeneity often occurs in randomized trials as well when there is non-compliance, which becomes problematic if it is caused by unobserved variables that are risk factors for the outcome. In that case, the usual regression estimators may not be consistent.

Instrumental variable method is an approach to yield unbiased estimate of the endogenous exposure. Although, the requirements of an IV depend on a particular analytic method one uses, the following three conditions are sufficient to define the IV (Brookhart et al. 2010): (i) an IV V has a causal effect on the exposure Z, (ii) V affects the outcome Y only through Z, (iii) V is unrelated to an unmeasured confounder U. In randomized trials, a randomization assignment indicator is often used as an IV to construct IV estimators for meaningful casual effects of treatment on the outcome (Robins and Tsiatis 1991, Loeys and Goetghebeur 2003, Loeys et al. 2005, Nie et al. 2011).

Structural linear equation models, which is one of the methods to represent causal effects, are considered to develop our IV method (Hernan and Robins 2006). The equation of interest (response model) relates Y to Z and a measured confounder W via a linear model. The other equation (exposure model) relates Z to V and W linearly. The regression parameters in the outcome model are identified using the instrument. For the case of no confounders, an IV estimator is given as the ratio of two covariance estimators, $\widehat{cov}(Y, V)/\widehat{cov}(Z, V)$. For the case where there are confounders, generalized method of moments (Hansen 1982) or two-stage least squares (Theil 1953) are used.

There has been considerable work with complete data, where both outcome and exposure are fully observed and assumed to satisfy semiparametric linear or nonlinear or nonparametric structural equation models with unspecified error distributions (Theil 1953, Amemiya 1974; 1982, Newey 1990, Chen and Portnoy 1996, Newey et al. 1999, Newey and Powell 2003). The popular two-stage least squares estimator has an explicit form, with a well-characterized sampling distribution and plug-in variance estimation, making inference straightforward (Bollen 1996, Bollen et al. 2007). However, if either response or exposure are incompletely observed, such semiparametric methods are not applicable. There has been limited work addressing two-stage IV estimation with such coarsening.

If we observe only a subset of the complete-data sample space where true data lie, then we refer to this kind of data as coarsened data (Heitjan and Rubin 1991). The various ways of coarseness include missing, rounding, heaping, right-censoring and so on. In this article, we focus on right-censored data, however, other types of coarsened data such as truncated data can be analyzed with our methods. Our methods do not cover coarsened data such as age heaping, whose valid inference is achieved by multiple imputation (Heitjan and Rubin 1990).

An important special case of coarsened data occurs with time to event outcomes and exposures, which may be subject to both censoring and truncation. For right censored data, there have been attempts to extend classical two-stage IV estimators. Robins and Tsiatis (1991) developed the IV estimators for correcting non-compliance in randomized trials using rank preserving structural failure time models. These models are an alternative to the usual two-stage models which are specially tailored to failure time data. A limitation is that the semiparametric estimation methods require that the censoring time is always known, as with fixed follow-up, and hence censoring due to dropout and other coarsening are not permitted. Bijwaard (2008) extended Robins and Tsiatis (1991) to generalized accelerated failure time models under similar censoring assumptions. Loeys and Goetghebeur (2003) proposed the IV estimators for the effect of treatment actually received in a randomized trial with all-or-nothing compliance based on the proportional hazard models. These methods were extended to allow more general exposure level and covariates to be included in the causal proportional hazard model (Loeys et al. 2005). Nie et al. (2011) proposed the IV estimators for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring, which are extensions of the methods of Baker (1998). Brännäs (2000) considered ad hoc two-stage estimators for the standard linear structural equation models which are IV adaptations of the symmetric trimmed least squares (Powell 1986a) and the Buckley-James (Buckley and James 1979) estimators for right censored data. However, the theoretical properties of these procedures were not investigated and a rigorous investigation of two stage IV estimation in linear models with right censoring is not apparent in the literature.

In Section 2, we propose a general framework for two-stage IV estimation of semiparametric linear structural equation models for outcome and exposure which accommodates coarsened data. The main requirement is that there exist semiparametric methods for fitting linear models to the outcome and exposure: this condition is satisfied under a wide range of censoring and truncation schemes commonly encountered in longitudinal studies. At stage 1, we construct and fit the reduced response and exposure models. At stage 2, we estimate the regression parameters in the true response model using a weighted minimum distance method based on the stage 1 results. This yields a closed form estimator, with a particular choice of weight leading to the standard two-stage least squares estimator with fully observed data. For the case of right censoring, the procedure does not require that the censoring time is always observed. We prove that our estimators are consistent and asymptotically normal, and provide a theoretically justified resampling technique for making inferences. The optimal weight is identified resulting in the minimum variance estimator, which may be superior to an usual two-stage estimator.

In Section 3, we discuss details related to the implementation of our semiparametric estimator when either response or exposure may be censored, employing existing estimators for accelerated failure time models under censoring. These methods are shown to perform well in simulations reported in Section 4, where naive estimation which ignores the unmeasured confounded may produce severely biased estimates of exposure effects. The practical utility of the methods is illustrated in a study of the comparative effectiveness of colon cancer treatments, where the effect of treatment on survival is confounded by patient health status.

2.2 General Framework for Coarsened Data

2.2.1 Model and estimation

For $i = 1, ..., n$, suppose that Y_i is the response variable, Z_i is the exposure variable, $V_i = (V_{i1}, ..., V_{ip})^T$ is the $p \times 1$ vector of the IVs, $W_i = (W_{i1}, ..., W_{iq})^T$ is the $q \times 1$ vector of the measured confounders, and U_i is the unmeasured confounder.

We consider the following linear response model,

$$\begin{aligned} Y_i &= \alpha_{yo} + \alpha_{yz} Z_i + \alpha_{yw}^T W_i + \alpha_{yu} U_i + \varepsilon_i \\ &= \alpha_{yo} + \alpha_0^T X_i + \varepsilon_i^*, \end{aligned} \tag{2.1}$$

where $\varepsilon_i^* = \alpha_{yu} U_i + \varepsilon_i$, $\alpha_{yw} = (\alpha_{yw,1}, ..., \alpha_{yw,q})^T$, $\alpha_0^T = (\alpha_{yz}, \alpha_{yw}^T)$, $X_i^T = (Z_i, W_i^T)$ and $E(\varepsilon_i \mid X_i, U_i) = 0$ by construction.

The linear model for the exposure is given by,

$$Z_i = \beta_{zo} + \beta_{zv}^T V_i + \beta_{zw}^T W_i + \beta_{zu} U_i + \delta_i, \tag{2.2}$$

where $\beta_{zv} = (\beta_{zv,1}, ..., \beta_{zv,p})^T$, $\beta_{zw} = (\beta_{zw,1}, ..., \beta_{zw,q})^T$ and $E(\delta_i \mid V_i, W_i, U_i) = 0$ by construction. One may rewrite model (2.2) as

$$Z_i = \beta_{zo} + \beta_0^T D_i + \delta_i^*, \tag{2.3}$$

where $\delta_i^* = \beta_{zu} U_i + \delta_i$, $\beta_0^T = (\beta_{zv}^T, \beta_{zw}^T)$, and $D_i^T = (V_i^T, W_i^T)$. We will call (3.27) the reduced exposure model.

The implied model for X_i is

$$
\begin{aligned}
X_i = \begin{pmatrix} Z_i \\ W_i \end{pmatrix} &= \begin{pmatrix} \beta_{zo} \\ 0_{q\times 1} \end{pmatrix} + \begin{pmatrix} \beta_{zv}^T & \beta_{zw}^T \\ 0_{q\times p} & I_q \end{pmatrix} \begin{pmatrix} V_i \\ W_i \end{pmatrix} + \begin{pmatrix} \delta_i^* \\ 0_{q\times p} \end{pmatrix} \\
&= \beta_{zo}^* + B_0^T D_i + \delta_i^{**}, \qquad (2.4)
\end{aligned}
$$

where B_0 is the $(p+q)\times(1+q)$ parameter matrix, $0_{q\times p}$ is a $q\times p$ zero matrix and I_q is a $q\times q$ identity matrix.

Substituting (3.28) into (3.26) gives

$$
Y_i = \gamma_{yo} + \gamma_0^T D_i + \tau_i, \qquad (2.5)
$$

where $\gamma_{yo} = \alpha_{yo} + \alpha_0^T \beta_{zo}^*$ is an intercept, $\gamma_0 = (\gamma_{yv}^T, \gamma_{yw}^T)^T = B_0\alpha_0$ is a $(p+q)\times 1$ parameter vector and $\tau_i = \varepsilon_i^* + \alpha_0^T \delta_i^{**}$. We will call (3.29) the reduced response model.

Two-stage IV estimation will be discussed based on the assumption that conditional on D_i, (τ_i, δ_i^*) is an independent and identically distributed sequence with mean zero and covariance matrix Σ_e. The simple sufficient condition to make the assumption of $E(\tau_i \mid D_i) = E(\delta_i^* \mid D_i) = 0$ hold is $E(\varepsilon_i \mid D_i) = E(\delta_i \mid D_i) = E(U_i \mid D_i) = 0$.

Remark 1. *From the assumption of exclusion restriction (Angrist et al. 1996) and the assumption of $E(\varepsilon_i \mid Z_i, W_i, U_i) = 0$ in the response model* (3.26)*, if follows that $E(\varepsilon_i \mid Z_i, V_i, W_i, U_i) = 0$. Clearly, $E(\varepsilon_i \mid Z_i, V_i, W_i, U_i) = 0$ implies $E(\varepsilon_i \mid D_i) = 0$. From the assumption of $E(\delta_i \mid D_i, U_i) = 0$ in the exposure model* (2.2)*, it follows that $E(\delta_i \mid D_i) = 0$. Thus we need $E(U_i \mid D_i) = 0$, which will be called IV independence assumption, to have $E(\tau_i \mid D_i) = E(\delta_i^* \mid D_i) = 0$. The IV independence assumption implies that the unmeasured confounder is balanced well between the strata generated by the instrument and the confounders.*

In coarsened data, rather than observing Y_i, we observe $\tilde{Y}_i = \psi(Y_i)$, where $\psi(\cdot)$ is some known function of Y_i. For example, in the setting of the accelerated failure time model, $\psi(Y_i) = \min(Y_i, C_i^Y)$, where Y_i is a log of failure time and C_i^Y is the corresponding log of censoring time.

The usual estimator for α_0 in (3.26) does not give a consistent estimator because X_i and ε_i^* are correlated with shared U_i, hence $E(\varepsilon^* \mid X_i)$ is not equal to zero in general unless $E(U_i \mid X_i) = E(U_i \mid Z_i, W_i^T) = 0$ for all $i = 1, ..., n$. However, since $E(\tau_i \mid D_i) = 0$ in the model (3.29), γ_0 can be consistently estimated using the data $\{(\tilde{Y}_1, D_1^T), ..., (\tilde{Y}_n, D_n^T)\}$. The proposed IV estimation method is developed under the condition that the estimators of $\theta_0^T = (\gamma_0^T, \beta_0^T)$ satisfying the below two assumptions exist.

Assumption 1. *The estimator $\hat{\theta}^T = (\hat{\gamma}^T, \hat{\beta}^T)$ converges in probability to $\theta_0^T = (\gamma_0^T, \beta_0^T)$.*

Assumption 2. *The random quantity $n^{1/2}(\hat{\theta} - \theta_0)$ converges in distribution to a mean 0 multivariate normal distribution with the covariance matrix Σ_{θ_0}.*

The covariance matrix Σ_{θ_0} consists of four block matrices,

$$\Sigma_{\theta_0} = \begin{pmatrix} \Sigma_{\gamma_0} & \Sigma_{\gamma_0,\beta_0} \\ \Sigma_{\beta_0,\gamma_0} & \Sigma_{\beta_0} \end{pmatrix}.$$

The estimator for B_0^T, $\hat{B}^T$, is defined as

$$\begin{pmatrix} \hat{\beta}_{zv}^T & \hat{\beta}_{zw}^T \\ 0_{q\times p} & I_q \end{pmatrix}.$$

Consistent and asymptotic normal estimators are obtained by the least squares for fully observed data and by the Buckley-James (Buckley and James 1979) and the rank (Prentice 1978) estimators for right-censored data. Asymptotical properties of the Buckley-James and rank estimators were studied by Ritov (1990), Tsiatis (1990), Lai and Ying (1991aa;a), Ying (1993) and Jin et al. (2006b).

Given the consistent estimators $\hat{\gamma}$ and $\hat{B}$, a consistent estimator for α_0 can be obtained by minimizing the weighted quadratic distance criterion

$$(\hat{\gamma} - \hat{B}\alpha_0)^T A_n (\hat{\gamma} - \hat{B}\alpha_0),$$

where A_n is a non-negative definite weighting (symmetric) matrix which may depend on the data, and $A_n/n = A + o_p(1)$. The minimum distance estimator (MDE) is given by

$$\hat{\alpha} = (\hat{B}^T A_n \hat{B})^{-1} \hat{B}^T A_n \hat{\gamma}.$$

For complete data, the two-stage least squares (2SLS) estimator is obtained by replacing the exposure by its predicted value calculated from fitting the reduced exposure model with the usual least squares. Define centered vectors as $X_{i(c)} = X_i - \bar{X}$ and $D_{i(c)} = D_i - \bar{D}$, where $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ and $\bar{D} = n^{-1}\sum_{i=1}^n D_i$. Let $X_{(c)}$ and $D_{(c)}$ be the matrices with the ith rows of $X_{i(c)}$ and $D_{i(c)}$. Then, the two-stage least squares estimator for α_0 can be written as

$$\hat{\alpha}_{2sls} = (\hat{X}_{(c)}^T \hat{X}_{(c)})^{-1} \hat{X}_{(c)}^T Y,$$

where $Y = (Y_1, ..., Y_n)^T$ and $\hat{X}_{(c)} = D_{(c)}\hat{B}$. From $\hat{X}_{(c)} = D_{(c)}\hat{B}$, it follows that

$$\hat{\alpha}_{2sls} = \left\{\hat{B}^T (D_{(c)}^T D_{(c)}) \hat{B}\right\}^{-1} \hat{B}^T D_{(c)}^T Y.$$

We can see that this is equivalent to $\hat{\alpha}$ with $A_n = D_{(c)}^T D_{(c)}$ and $\hat{\gamma} = (D_{(c)}^T D_{(c)})^{-1} D_{(c)}^T Y$. Therefore, $\hat{\alpha}$ contains the two-stage least squares estimator as a special case.

Next, we present the major theoretical results for the proposed IV estimator.

Theorem 1. *Under Assumption 1, $\hat{\alpha}$ converges in probability to α_0.*

Proof. $\hat{\alpha} = \left(\hat{B}^T (A_n/n) \hat{B}\right)^{-1} \hat{B}^T (A_n/n)\hat{\gamma} = (B_0^T A B_0)^{-1} B_0^T A \gamma_0 + o_p(1) = \alpha_0 + o_p(1).$ □

Theorem 2. *Under Assumptions 1 and 2, $n^{1/2}(\hat{\alpha} - \alpha_0)$ converges in distribution to a mean 0 multivariate normal distribution with the covariance matrix*
$(B_0^T A B_0)^{-1} B_0^T A \Omega(\alpha_0) A B_0 (B_0^T A B_0)^{-1}$, *where* $\Omega(\alpha_0) = var\left\{n^{1/2}(\hat{\gamma} - \hat{B}\alpha_0)\right\}$.

Proof.

$$
\begin{aligned}
n^{\frac{1}{2}}(\hat{\alpha} - \alpha_0) &= n^{\frac{1}{2}}\left\{(\hat{B}^T A_n \hat{B})^{-1}\hat{B}^T A_n \hat{\gamma} - (\hat{B}^T A_n \hat{B})^{-1}\hat{B}^T A_n \hat{B}\alpha_0\right\} \\
&= (\hat{B}^T A_n \hat{B})^{-1}\hat{B}^T A_n n^{\frac{1}{2}}(\hat{\gamma} - \hat{B}\alpha_0).
\end{aligned}
$$

By multivariate Slutsky theorem, Theorem 2 holds due to the facts that $(\hat{B}^T A_n \hat{B})^{-1}\hat{B}^T A_n = (B_0^T A B_0)^{-1} B_0^T A + o_p(1)$ and that $n^{1/2}(\hat{\gamma} - \hat{B}\alpha_0)$ converges to a mean 0 multivariate normal distribution with the covariance matrix $\Omega(\alpha_0)$. □

Remark 2. *Although Theorems 1 and 2 look straightforward, those theorems are very useful because they convert the problem of finding consistent and asymptotically normal IV estimators to that of finding well established estimators such as rank estimators for right-censored data and Powell's estimators (Powell 1984; 1986b) for truncated data. We will present four Corollaries to state the asymptotic properties of the proposed IV estimators for four types of data and those Corollaries are directly followed by Theorems 1 and 2.*

The lower bound of the above covariance matrix of $n^{1/2}(\hat{\alpha} - \alpha_0)$ is $(B_0^T \Omega(\alpha_0)^{-1} B_0)^{-1}$. This is obtained by taking $A = \Omega(\alpha_0)^{-1}$. The corresponding $\hat{\alpha}$ is obtained by using the weight $A_n = \hat{\Omega}(\hat{\alpha})^{-1}$, which is a consistent estimator for $\Omega(\alpha_0)^{-1}$ if $\hat{\alpha}$ is consistent for α_0. In order to compute $A_n = \hat{\Omega}(\hat{\alpha})^{-1}$, we need a consistent estimator for α_0. In practice, we may use the following initial estimator, $\hat{\alpha}_I = (\hat{B}^T \hat{B})^{-1}\hat{B}^T \hat{\gamma}$, with with identity weight matrix $A_n = I$.

The matrix $\Omega(\alpha_0)$ depends on the asymptotic covariance matrix of $(\hat{\gamma}^T, \hat{\beta}^T)$. Note that

$$
\hat{\gamma} - \hat{B}\alpha_0 = g(\hat{\gamma}, \hat{\beta}) = \begin{pmatrix} \hat{\gamma}_{yv} - \alpha_{yz}\hat{\beta}_{zv} \\ \hat{\gamma}_{yw} - \alpha_{yz}\hat{\beta}_{zw} - \alpha_{yw} \end{pmatrix}, \ \dot{g}(\hat{\gamma}, \hat{\beta}) = \begin{pmatrix} I_{p+q} \\ -\alpha_{yz} I_{p+q} \end{pmatrix},
$$

where $\dot{g}(\theta)$ is the first derivative of $g(\theta)$ with respect to θ. Thus $\Omega(\alpha_0) = \Sigma_{\gamma_0} - \alpha_{yz}(\Sigma_{\gamma_0,\beta_0} + \Sigma_{\beta_0,\gamma_0}) + \alpha_{yz}^2 \Sigma_{\beta_0}$. For the 2SLS estimator in complete data, $\Omega(\alpha_0) = \text{var}(\varepsilon_i^* - \alpha_{yz}\delta_i^*)\bar{M}^{-1}$, where $\bar{M} = \lim_{n\to\infty} D_{(c)}^T D_{(c)}/n$.

If we have one IV $(p = 1)$ and B_0 is nonsingular, then $\alpha_0 = B_0^{-1}\gamma_0$ and $\hat{\alpha} = \hat{B}^{-1}\hat{\gamma}$. In this set-up, $\hat{\alpha}$ does not dependent on the weight matrix A_n. The covariance matrix of $n^{1/2}(\hat{\alpha} - \alpha_0)$ with one IV is given by $(B_0^T \Omega(\alpha_0)^{-1} B_0)^{-1}$ which is the lower bound of the covariance matrix of $n^{1/2}(\hat{\alpha} - \alpha_0)$. If there are no confounders, then $\hat{\alpha} = \hat{\gamma}_{yv}/\hat{\beta}_{zv}$, which is just the ratio of the two regression parameter estimators.

2.2.2 A resampling method for variance estimation

In our setting, we are interested in drawing inferences for parameters, say β, under semiparametric models. One may use the estimating equation because the resulting solution is consistent and asymptotically normal under mild conditions. The estimator $\hat{\beta}$ for β_0 can be easily computed by solving the corresponding estimating function. However, the variance of $\hat{\beta}$ can involve complicated nonparametric function estimation if the estimating equation is not smooth enough in β. For example, the variance of rank-based estimator for the accelerated failure time model contains the derivative of hazard function of the error terms. Thus direct computation of the variance would require nonparametric density estimation. To avoid this difficulty, resampling methods can be used.

Jin et al. (2001) proposed a resampling method by perturbing an objective function. If the objective function has its first derivative, i.e. an estimating equation, then it is equivalent to perturb the estimating equation. The method of Jin et al. (2001) provides a valid inference procedure under the assumption that both the estimating equation and its perturbed one have 'good' quadratic equations around the true value of the parameter. That assumption holds in a wide range of regression problems including the estimations based on L_p norm and Wilcoxon statistic, and truncated median regression under mild conditions. Thus, dealing with data coarsening such as censoring and truncation, the resampling method of Jin et al. (2001) can be effectively used to perform variance estimation of the proposed IV estimators. For right-censored data, the method of Jin et al. (2001) has been extended to rank estimation (Jin et al. 2003; 2006a), Buckley-James estimation (Jin et al. 2006b) and local Buckley-James estimation (Pang et al. 2014) of the accelerated failure time model. Details about the inference using the resampling are provided in the next section.

2.3 Inference

We start with sketching our two-stage IV method which involves solving two separate estimating equations. To obtain $\hat{\gamma}$ and $\hat{\beta}$, we find the roots of the estimating functions,

$$U_1(\gamma) = \sum_{i=1}^{n} U_{1i}(\gamma),\ U_2(\beta) = \sum_{i=1}^{n} U_{2i}(\beta), \tag{2.6}$$

where $U_1(\gamma)$ and $U_2(\beta)$ are the estimating equations for the response and exposure reduced models, (3.29) and (3.27), respectively.

Throughout the rest of the paper, we assume that the response is right-censored. Then, we mainly consider two cases: Case 1, the observed exposure is continuous; Case 2, the observed exposure is binary. We further divide each case into the two sub-cases: Case 1-1, the exposure is fully observed; Case 1-2, the exposure is right or left-censored; Case 2-1, the binary exposure is observed via coarsening of the latent exposure; Case 2-2, the binary exposure is directly observed.

For all of the cases except for Case 2-2, the equation $U_1(\gamma)$ is the Gehan estimating equation for the accelerated failure time model (Fygenson and Ritov 1994, Jin et al. 2003). The equation $U_2(\beta)$ is the normal equation for the linear model for Case 1-1, the Gehan estimating equation for Case 1-2 and the probit score equation for Case 2-1. For Case 2-2, the equation $U_1(\gamma)$ is the normal equation for the local Buckley-James estimator of heteroscedastic accelerated failure time model (Pang et al. 2014) and $U_2(\gamma)$ is the normal equation for the linear probability model. The estimators, $\hat{\gamma}$ and $\hat{\beta}$, are solutions of those estimating equations. They are consistent for

γ_0 and β_0 and asymptotically joint normal under certain conditions (see regularity conditions in the Appendix). These asymptotic properties of rank estimators follow from Tsiatis (1990) and Ying (1993) and those of local Buckley-James estimator follow from Pang et al. (2014), while for the least squares and maximum likelihood estimators in Case 1-1, 2-1 and 2-2, the results are standard. By Theorems 1 and 2, the two-stage IV estimator, $\hat{\alpha}$, is consistent for $\boldsymbol{\alpha}$ and asymptotically normal.

To estimate the variance of the two-stage IV estimator, we generate the joint distribution of $\hat{\gamma}$ and $\hat{\beta}$ by perturbing the two estimating equations with the same positive random variables whose mean and variance are one and which are independent of the data $(\tilde{Y}_i, Z_i, D_i)(i = 1, ..., n)$. Let $R = (R_1, ..., R_n)$ be the vector of random variables used for perturbation. The perturbed estimating equations are given by

$$U_1^*(\gamma) = \sum_{i=1}^{n} U_{1i}^*(\gamma) = \sum_{i=1}^{n} U_{1i}(\gamma) R_i,\ U_2^*(\beta) = \sum_{i=1}^{n} U_{2i}^*(\beta) = \sum_{i=1}^{n} U_{2i}(\beta) R_i. \tag{2.7}$$

We perturbed the two estimating equations by multiplying the original estimating equations by the same $R_i(i = 1, ..., n)$, which ensures that the covariance of the estimating equations is correctly accounted for in the resampling. For $l = 1, 2$, under mild conditions, $n^{-1/2} U_l^*(\cdot)$ has mean 0 and approximately the same variance as $n^{-1/2} U_l(\cdot)$ conditionally on the data (Jin et al. 2003). In addition, the conditional covariance matrix of $n^{-1/2} U_1^*(\gamma)$ and $n^{-1/2} U_2^*(\beta)$ given the data converges to the asymptotic covariance matrix of $n^{-1/2} U_1(\gamma)$ and $n^{-1/2} U_2(\beta)$. Based on those arguments, we can obtain the joint distribution of $\hat{\gamma}$ and $\hat{\beta}$. For accelerated failure time model, the resampling method used in (2.7) is sufficient to generate marginal distribution of $\hat{\gamma}$ or $\hat{\beta}$ (Jin et al. 2003). However, to generate the joint distribution of the estimators, we need to modify (2.7), as discussed below. The resampling of local Buckley-James estimator is similar to that of rank estimator, but is more complex because perturbing the Kaplan-Meier estimator of the error distribution is required.

Suppose we repeatedly perturb the estimating equations a large number of times, say K, while fixing the data, with the kth resampled perturbing random variables denoted as $R^k = (R_1^k, ..., R_n^k)$. Denote by $\text{tr}(\hat{\gamma}^k) = \{\text{tr}(\hat{\gamma}_{yv}^k), \text{tr}(\hat{\gamma}_{yw}^k)\}$ and $\text{tr}(\hat{\beta}^k) = \{\text{tr}(\hat{\beta}_{zv}^k), \text{tr}(\hat{\beta}_{zw}^k)\}$ the solutions of the kth perturbed estimating equations. Then we can construct the kth resampled $\hat{\alpha}$, $\hat{\alpha}^k$, from $\hat{\gamma}^k$ and $\hat{\beta}^k$,

$$\hat{\alpha}^k = \{\text{tr}(\hat{B}^k) A_n^* \hat{B}^k\}^{-1} \text{tr}(\hat{B}^k) A_n^* \hat{\gamma}^k,$$

where $\hat{B}^k$ is defined as

$$\begin{pmatrix} \text{tr}(\hat{\beta}_{zv}^k) & \text{tr}(\hat{\beta}_{zw}^k) \\ 0_{q\times p} & I_q \end{pmatrix}.$$

The weight matrix, A_n^*, is defined as the inverse of the empirical covariance matrix of $\{n^{1/2}(\hat{\gamma}^1 - \hat{B}^1\hat{\alpha}_I), ..., n^{1/2}(\hat{\gamma}^K - \hat{B}^K\hat{\alpha}_I)\}$ and $\hat{\alpha}_I = (\hat{B}^T\hat{B})^{-1}\hat{B}^T\hat{\gamma}$. The distribution of $\hat{\alpha}$ can be approximated by the empirical distribution of $\{\hat{\alpha}^1, ..., \hat{\alpha}^K\}$.

2.3.1 Case 1-1: Fully-observed continuous exposure

Here we specify the methods for Case 1-1. We employ the AFT model for the censored response model (3.29), assuming that $(\tau_1, ..., \tau_n)$ are independent error terms with a common, but unspecified distribution. The response vector Y is the vector of log of survival times. Let $C^Y = (C_1^Y, ..., C_n^Y)^T$ be the vector of log of censoring times for Y. Assume that Y_i and C_i^Y are independent conditionally on $D_i^T = (V_i^T, W_i^T)$ and C_i^Y is not affected by U_i. The data consists of $(\tilde{Y}_i, \Delta_i^Y, D_i)$, where $\tilde{Y}_i = \min(Y_i, C_i^Y)$, $\Delta_i^Y = I(Y_i \le C_i^Y)$. Here, $I(Q)$ is one when a statement Q is true, and zero otherwise.

Define $e_i(\gamma) = \tilde{Y}_i - \gamma^T D_i$, $N_i(\gamma; t) = \Delta_i^Y I\{e_i(\gamma) \le t\}$ and $Y_i(\gamma; t) = I\{e_i(\gamma) \ge t\}$. Note $N_i(\gamma; t)$ and $Y_i(\gamma; t)$ are the counting process and at-risk process on the residual time scale.

Write

$$S^{(0)}(\gamma;t) = n^{-1}\sum_{i=1}^{n} Y_i(\gamma;t),\ S^{(1)}(\gamma;t) = n^{-1}\sum_{i=1}^{n} Y_i(\gamma;t)D_i.$$

The Gehan-type rank estimator $\hat{\gamma}_G$ is a root of the following estimating equation.

$$U_{1,G}(\gamma) = \sum_{i=1}^{n}\int_{-\infty}^{\infty} S^{(0)}(\gamma;t)\{D_i - \bar{D}(\gamma;t)\}dN_i(\gamma;t), \tag{2.8}$$

where $\bar{D}(\gamma;t) = S^{(1)}(\gamma;t)/S^{(0)}(\gamma;t)$. Or equivalently,

$$U_{1,G}(\gamma) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i^y(D_i - D_j)I\{e_i(\gamma) \le e_j(\gamma)\}. \tag{2.9}$$

The above equation is monotone in each component of γ (Fygenson and Ritov 1994).

We can generate the resampled rank estimators by solving the following perturbed estimating equation,

$$U_{1,G}^{*}(\gamma) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i^y(D_i - D_j)I\{e_i(\gamma) \le e_j(\gamma)\}R_iR_j, \tag{2.10}$$

where $R_i(i = 1, ..., n)$ are positive random variables with $E(R_i) = var(R_i) = 1$, and independent of the data. The perturbation in (2.10) is more complex than the usual approach, in which each term in the estimating equation is multiplied by a single R_i. Jin et al. (2006a) showed that the resampling technique in (2.10) can produce joint distribution of the rank estimators from separate marginal linear models. Since the reduced response model is the accelerated failure time model, the resampling technique in (2.10) is essential to generate joint distribution of the estimators.

For the exposure which is fully observed, one may use least squares estimation. For simplicity, we assume that the error terms, $(\delta_1^*, ..., \delta_n^*)$, in the reduced exposure model are independent with a common unspecified distribution. The least squares estimator for β_0, which will be denoted as

$\hat{\beta}_L$, is obtained by solving the following estimating equation, which is the normal equation,

$$U_{2,L}(\beta) = \sum_{i=1}^{n}(D_i - \bar{D})(Z_i - D_i^T\beta). \tag{2.11}$$

We can generate the resampled least squares estimators by solving the following perturbed estimating equation,

$$U^*_{2,L}(\beta) = \sum_{i=1}^{n}(D_i - \bar{D})(Z_i - D_i^T\beta)R_i, \tag{2.12}$$

where $R_i(i = 1, ..., n)$ are the same random variables used in (2.10). Employing the same perturbations is essential to generating the joint distribution of $(\hat{\gamma}_G, \hat{\beta}_L)$.

Below we present corollaries and a theorem for the asymptotic properties of the two-stage IV estimator with the Gehan rank estimator and the least square estimator, and the approximation of the asymptotic distribution of the two-stage IV estimator via the above resampling.

Corollary 1. *For Case 1-2, the Gehan rank estimator for γ_0, denoted as $\hat{\gamma}_G$, and the lest squares estimator for β_0, denoted as $\hat{\beta}_L$, satisfy Assumption 1 and 2 under the conditions A1-A4 in the appendix. Therefore the two-stage estimator, $\hat{\alpha}$, with $\hat{\gamma}_G$ and $\hat{\beta}_L$ converges in probability to α_0 and asymptotically normal by Theorem 1 and 2.*

Theorem 3. *For Case 1-2, under the conditions A1-A4 in the appendix, the asymptotic distribution of $\hat{\alpha}$ can be estimated by the empirical distribution of $\hat{\alpha}^* = (\hat{\alpha}^1, ..., \hat{\alpha}^K)$ conditionally on the data, where $\hat{\alpha}^k(k = 1, ..., K)$ is the resampled $\hat{\alpha}$ at the kth perturbation.*

2.3.2 Case 1-2: Right or left-censored exposure

Next we specify the methods for Case 1-2 where both the response and exposure is right or left-censored. The model with left-censored data can be estimated by taking negative values of the parameter estimates obtained from the estimation method for right-censored data, where the outcome is multiplied by a negative sign. Left-censoring often occurs in the studies where the measurement of biomarkers or environmental substances is subject to detection limit. As Wang and Feng (2012) stated, the methods for regression with the covariates missing at random is not applicable to this case as the censoring of the covariates reflects the size of true value.

Akritas et al. (1995) developed the Theil-Sen estimator for the slope in a simple regression when both the response and exposure are possibly left-censored. They referred this to doubly censored data. The motivation of their methods comes from astronomical data where nondetections occur when observed values of the samples are below a certain level. Thus these nondetections become left-censored data points. Wang and Feng (2012) proposed multiple imputation for M-regression with left-censored covariates. Their method uses a linear quantile regression model to impute the censored values given the observed data. Bernhardt et al. (2014) proposed imputation method for left-censored covariates under accelerated failure time model. They use seminonparametric distribution to model the error term and assume the distribution of censored covariates conditional on observed variables is known. Thus their method relies on parametric models and the regression parameters are estimated by maximum likelihood estimation. The proposed method for Case 1-2 requires the presence of IVs which satisfy the three conditions mentioned in Introduction. Our method has been developed under the situation where there are unmeasured confounders, however, also can be applicable to the data without unmeasured confounders.

Our method is similar to that of Wang and Feng (2012) in the sense that model treating the exposure as the dependent variable is fitted to account for the censoring for the exposure. However, unlike the method of Wang and Feng (2012), our method does not involve imputation and censoring for the response is allowed, and can account for potential unmeasured confounders. In addition, Wang and Feng (2012) requires exogenous censoring, which may not hold if unmeasured confounders are present. Like our method, the method of Akritas et al. (1995) allows for doubly-censored data. However, their method cannot include other covariates in the model and assumes that the single covariate is independent of the error term. Bernhardt et al. (2014) considers censored covariates problem in the accelerated failure time model, but their method requires parametric assumptions on the error term and on the conditional distribution of censored covariates given the observed data.

The real example where our method can be applied is the study of Smith et al. (2005) who investigated the influence of C-reactive protein levels on blood pressure. They employed IV estimation using Mendelian randomization to account for unmeasured confounders with an IV being a specific gene relating with C-reactive protein. However, some measured values of C-reactive protein levels were left-censored because of limit of detections and those censored data was omitted from their analysis. Our method can be used to account for left-censoring and may give more reliable estimates.

To derive the asymptotic properties of the two-stage IV estimator for doubly-censored data, we directly use the theoretical results of Section 2 in Jin et al. (2006a), which discusses the rank regression for multivariate failure time data based on marginal accelerated failure time models. In Case 1-2, both reduced response and exposure models are the accelerated failure time models and we use the Gehan rank estimators to estimate the regression parameters. We will use Theorem 1 of Jin et al. (2006a) to derive consistency and asymptotic joint normality of the rank estimators obtained from the reduced models.

The rank estimation for both reduced models is conducted exactly in the same way as in Case 1-1. For completeness, we describe rank estimation for the reduced exposure model in detail. We consider the accelerated failure time model for the reduced exposure model (3.27) with assuming that $(\delta_1^*, ..., \delta_n^*)$ are independent error terms with a common, but unspecified distribution. Without loss of generality, let $Z = (Z_1, ..., Z_n)^T$ and $C^Z = (C_1^Z, ..., C_n^Z)$ be the vectors of the log of the exposure and the corresponding log of the censoring time. Assume that Z_i and C_i^Z are independent conditionally on $D_i^T = (V_i^T, W_i^T)$ and C_i^Z is not affected by U_i. The data consists of $(\tilde{Z}_i, \Delta_i^Z, D_i)$, where $\tilde{Z}_i = \min(Z_i, C_i^Z)$, $\Delta_i^Z = I(Z_i \leq C_i^Z)$.

Define $e_i(\beta) = \tilde{Z}_i - \beta^T D_i$, $N_i(\beta; t) = \Delta_i^Z I\{e_i(\beta) \leq t\}$ and $Y_i(\beta; t) = I\{e_i(\beta) \geq t\}$. Write

$$S^{(0)}(\beta; t) = n^{-1} \sum_{i=1}^{n} Y_i(\beta; t),\ S^{(1)}(\beta; t) = n^{-1} \sum_{i=1}^{n} Y_i(\beta; t) D_i.$$

The Gehan-type rank estimator $\hat{\beta}_G$ is a root of the following estimating equation.

$$U_{2,G}(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} S^{(0)}(\beta;t)\{D_i - \bar{D}(\beta;t)\} dN_i(\beta;t), \tag{2.13}$$

where $\bar{D}(\beta;t) = S^{(1)}(\beta;t)/S^{(0)}(\beta;t)$. Or equivalently,

$$U_{2,G}(\beta) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i^Z (D_i - D_j) I\{e_i(\beta) \le e_j(\beta)\}. \tag{2.14}$$

The above equation is monotone in each component of β.

We can generate the resampled rank estimators by solving the following perturbed estimating equation,

$$U_{2,G}^*(\beta) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i^Z (D_i - D_j) I\{e_i(\beta) \le e_j(\beta)\} R_i R_j, \tag{2.15}$$

where $R_i (i = 1, ..., n)$ are the same random variables used for perturbing the reduced response model.

Corollary 2. *For Case 1-2, the Gehan rank estimators for γ_0 and β_0, denoted as $\hat{\gamma}_G$ and $\hat{\beta}_G$, satisfy Assumption 1 and 2 under the conditions A1-A4 in the appendix. Therefore the two-stage estimator, $\hat{\alpha}$, with $\hat{\gamma}_G$ and $\hat{\beta}_G$ converges in probability to α_0 and asymptotically normal by Theorem 1 and 2.*

Theorem 4. *For Case 1-2, under the conditions A1-A4 in the appendix, the asymptotic distribution of $\hat{\alpha}$ can be estimated by the empirical distribution of $\hat{\alpha}^* = (\hat{\alpha}^1, ..., \hat{\alpha}^K)$ conditionally on the data.*

2.3.3 Case 2-1: Dichotomized exposure via coarsening of latent exposure

In Case 2-1, we assume that Z_i in the true response model (3.26) is a latent variables which are not directly observed. The observed exposure is denoted as one when the latent exposure variable, Z_i, is greater than zero and as zero otherwise. That is, the observed exposure is $\tilde{Z}_i = I(Z_i > 0)$. This kind of approach using latent variables is often employed to model dummy variables (Heckman 1978). Thus, in Case 2-1 analysis, we measure the effect of the latent exposure on the response. Through this analysis, we can test whether their is a causal effect of the exposure on the response, however, cannot estimate binary effect directly unlike Case 2-2 analysis, which will be discussed next subsection. An additional analysis for Case 2-1 is to convert the estimate of the latent exposure to that of binary exposure. However, in the appendix, we showed that this conversion dose not give a correct binary estimate in general.

We fit the probit model to the observed binary exposure, which is a coarsening of the underlying latent variable. For identification of the model parameters, we assume that $\delta_i^*(i = 1, ..., n)$ independently follows a standard normal distribution. The probit model is

$$P(\tilde{Z}_i = 1) = P(Z_i > 0) = \Phi(\beta_{zo} + \beta_0^T D_i),$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard normal random variable.

The maximum likelihood estimator for β_0, $\hat{\beta}_M$, is obtained by solving the following estimating equation, which is a likelihood score equation,

$$U_{2,M}(\beta_o, \beta) = \sum_{i=1}^{n} \frac{\{\tilde{Z}_i - \Phi(\beta_o + \beta^T D_i)\}\phi(\beta_o + \beta^T D_i)}{\Phi(\beta_o + \beta^T D_i)\{1 - \Phi(\beta_o + \beta^T D_i)\}} D_i,$$

where β_o is a parameter for an intercept and $\phi(\cdot)$ is the density function of standard normal random variable.

To generate the resampled maximum likelihood estimator for β_0, we solve the perturbed score equation,

$$U_{2,M}^*(\beta_o, \beta) = \sum_{i=1}^{n} \frac{\{\tilde{Z}_i - \Phi(\beta_o + \beta^T D_i)\}\phi(\beta_o + \beta^T D_i)}{\Phi(\beta_o + \beta^T D_i)\{1 - \Phi(\beta_o + \beta^T D_i)\}} D_i R_i, \tag{2.16}$$

where $R_i(i = 1, ..., n)$ are also used to perturb the reduced response model.

Corollary 3. *For the data with the censored response and dichotomized exposure (Case 3), the Gehan rank estimators for γ_0, denoted as $\hat{\gamma}_G$, and the maximum likelihood estimator for β_0, denoted as $\hat{\beta}_M$, satisfy Assumption 1 and 2 under the conditions A1-A4 in the appendix.*

Therefore the two-stage estimator, $\hat{\alpha}$, with $\hat{\gamma}_G$ and $\hat{\beta}_M$ converges in probability to α_0 and asymptotically normal by Theorem 1 and 2.

Theorem 5. *For Case 3, under the conditions A1-A4 in the appendix, the asymptotic distribution of $\hat{\alpha}$ can be estimated by the empirical distribution of $\hat{\alpha}^* = (\hat{\alpha}^1, ..., \hat{\alpha}^K)$, conditionally on the data.*

2.3.4 Case 2-2: Binary exposure

In Case 2-2, we use the binary exposure itself and does not consider coarsening of the exposure. That is, Z_i in the true response model (3.26) is binary. Thus, we estimate the causal effect of binary exposure on the response directly. In this case, the exposure model (2.2) becomes a linear probability model and the variance of the error terms depends on the covariates.

Recall that the true exposure model and the corresponding reduced exposure model are given by

$$\begin{aligned} Z_i &= \beta_{zo} + \beta_{zv}^T V_i + \beta_{zw}^T W_i + \beta_{zu} U_i + \delta_i, \\ &= \beta_{zo} + \beta_0^T D_i + \delta_i^*, \end{aligned} \tag{2.17}$$

where $E(\delta_i \mid D_i, U_i) = 0$ and $\text{var}(\delta_i \mid D_i, U_i) = \mu_z(D_i, U_i)(1 - \mu_z(D_i, U_i))$ by construction. As in Section 2·1, the reduced response model is given by

$$Y_i = \gamma_{yo} + \gamma_0^T D_i + \tau_i,$$

where $\tau_i = \varepsilon_i^* + \alpha_{yz}\delta_i^*$. Now the variance of τ_i depends on D_i as so does that of δ_i^*. Two-stage IV estimation for binary exposure will be discussed with the assumptions that $E(\tau_i \mid D_i) = E(\delta \mid D_i) = 0$. As in the previous cases, $E(\tau_i \mid D_i) = E(\delta_i^* \mid D_i) = 0$, holds if $E(\varepsilon_i \mid D_i) = E(\delta_i \mid D_i) = E(U_i \mid D_i) = 0$. Thus, an additional required condition is $E(U_i \mid D_i) = 0$.

In Remark 3, we show that with the IV independence assumption the reduced exposure model is still a linear probability model with the variance $\text{var}(\delta_i^* \mid D_i) = \mu_z^*(D_i)(1 - \mu_z(D_i))$ where

$\mu_z^*(D_i) = E(Z_i \mid D_i) = \beta_{zo} + \beta_0^T D_i$.

Remark 3. *By a simple probability argument,* $var(\delta_i^* \mid D_i) = E\{var(\delta_i^* \mid D_i, U_i) \mid D_i\} + var\{E(\delta_i^* \mid D_i, U_i) \mid D_i\}$ *and that* $var(\delta_i^* \mid D_i, U_i) = var(\delta_i \mid D_i, U_i)$*. From* $\mu_z(D_i, U_i) = \mu_z^*(D_i) + \beta_{zu} U_i$*, it follows that*

$$\begin{aligned} E\{var(\delta_i^* \mid D_i, U_i) \mid D_i\} &= E\{var(\delta_i \mid D_i, U_i) \mid D_i\} \\ &= \mu_z^*(D_i)(1 - \mu_z^*(D_i)) - \beta_{zu}^2 E(U_i^2 \mid D_i) \\ &+ \beta_{zu} E(U_i \mid D_i) - 2\beta_{zu}\mu_z^*(D_i) E(U_i \mid D_i). \end{aligned}$$

Note that $var\{E(\delta_i^* \mid D_i, U_i) \mid D_i\} = \beta_{zu}^2 var(U_i \mid D_i)$*. Since* $E(U_i \mid D_i) = 0$*,* $var(\delta_i^* \mid D_i) = \mu_z^*(D_i)(1 - \mu_z^*(D_i))$*.*

Since the reduced response model has a heteroscedastic error variance, we cannot use rank estimators to estimate the parameters. Instead, we use the local Buckley-James estimator (Pang et al. 2014) to account for heteroscedastic error variance. The authors showed that the local Buckley-James estimator is consistent and asymptotically normal under some regularity conditions. As in rank estimation, we assume that Y_i and C_i^Y are independent conditionally on D_i and C_i^Y is not affected by U_i.

The local Buckley-James estimator (Pang et al. 2014) was developed under the following model,

$$Y_i = \gamma_{yo} + \gamma_0^T D_i + \sigma(\gamma_0^T D_i)\omega_i, \tag{2.18}$$

where $\omega_i (i = 1, ..., n)$ are independent and identically distributed random variables with mean zero and standard deviation one. The function $\sigma(\gamma_0^T D_i)$ is a nonparametric function of $\gamma_0^T D_i$ and describes the heteroscedastic error variance which depends on $\gamma_0^T D_i$. The model (2.18) implies that the conditional variance of Y_i depends on the covariate D_i through $\gamma_0^T D_i$.

The model we may have is slightly different from the model (2.18),

$$Y_i = \gamma_{yo} + \gamma_0^T D_i + \sigma(\beta_0^T D_i)\omega_i, \tag{2.19}$$

where $\tau_i = \sigma(\beta_0^T D_i)\omega_i$ and $\omega_i (i = 1, ..., n)$ are independent and identically distributed random variables with mean zero and standard deviation one. Thus the conditional error variance in (2.19) is the function of $\beta_0^T D_i$, which is the mean function of the reduced exposure model. This comes from the fact that heteroscedastic error variance in the reduced response model is induced by that in the reduced exposure model. The local Buckley-James estimation will be performed to account for $\sigma(\beta_0^T D_i)$. To have the model (2.19), we need the condition that var$(\varepsilon_i^* \mid D_i)$ and cov$(\varepsilon_i^*, \delta_i^* \mid D_i)$ do not depend on D_i, which may hold when the conditional covariance matrix of $(\varepsilon_i, \delta_i, U_i)$ given D_i is fixed. Under this condition, var$(\tau_i \mid D_i)$ = constant + $\alpha_{yz}^2 \mu_z^*(D_i)(1 - \mu_z^*(D_i))$, hence we can write $\tau_i = \sigma(\beta_0^T D_i)\omega_i$.

Now we describe the procedure of the local Buckley-James estimation for the model (2.19). As in the Buckley-James estimation, we impute censored data by its estimated conditional mean,

$$\begin{aligned} E(Y_i \mid Y_i \geq C_i^Y, \tilde{Y}_i, D_i) &= E(e_i \mid Y_i \geq C_i^Y, \tilde{Y}_i, \beta_0^T D_i) + \gamma_0^T D_i \\ &= \frac{\int_{\tilde{Y}_i - \gamma_0^T D_i}^{\infty} u \, dF_{\theta_0}(u \mid \beta_0^T D_i)}{1 - F_{\theta_0}(\tilde{Y}_i - \gamma_0^T D_i \mid \beta_0^T D_i)} + \gamma_0^T D_i, \end{aligned}$$

where $\theta_0^T = (\gamma_0^T, \beta_0^T)$ and $F_{\theta_0}(u \mid \nu)$ is the unknown cumulative distribution function of the residual $e_i \equiv \tau_i = Y_i - \gamma_0^T D_i$ conditional on $\beta_0^T D_i = \nu$. Since $F_{\theta_0}(u \mid \beta_0^T D_i)$ depends on $\beta_0^T D_i$, it cannot be consistently estimated by the Kaplan-Meier estimate. Instead, we will use local Kaplan-Meier estimator (Dabrowska, 1987) to estimate $F_{\theta_0}(u \mid \beta_0^T D_i)$. The local Buckely-James estimation for the model (2.19) is performed as follows.

Step 1. Obtain an initial estimator for γ_0, for example, the Buckley-James estimator or the rank estimator.

Step 2. At the ath iteration, compute the imputed response value Y_i by

$$\hat{Y}_i(\gamma_a) = \Delta_i^Y \tilde{Y}_i + (1 - \Delta_i^Y)\hat{E}(Y_i \mid Y_i \geq C_i^Y, \tilde{Y}_i, \hat{\beta}_P^T D_i),\ i = 1, ..., n,$$

where $\hat{\beta}_P$ is the least squares estimator for β_0 in the linear probability model (2.17) and

$$\hat{E}(Y_i \mid Y_i \geq C_i^Y, \tilde{Y}_i, \hat{\beta}_P^T D_i) = \gamma_a^T D_i + \frac{\int_{e_i(\gamma_a)}^{\infty} u \, d\hat{F}_{\check{\theta}_a}(u \mid \hat{\beta}_P^T D_i)}{1 - \hat{F}_{\check{\theta}_a}\{e_i(\gamma_a) \mid \hat{\beta}_P^T D_i\}},$$

where $\check{\theta}_a^T = (\gamma_a^T, \hat{\beta}_P^T)$ and $e(\gamma_a) = \tilde{Y}_i - \gamma_a^T D_i$. The local Kaplan-Meier estimate of $F_\theta(t \mid \beta^T D_i)$ is obtained as follows.

$$\hat{F}_\theta(t \mid \beta^T D_i) = 1 - \prod_{j: e_j(\gamma) < t}^{n} \left\{ 1 - \frac{B_{nj}(\beta^T D_j)\Delta_i^Y}{\sum_{k=1}^n I\{e_k(\gamma) \geq e_j(\gamma)\} B_{nk}(\beta^T D_i)} \right\},$$

where $B_{nk}(\cdot)$, $k = 1, ..., n$, is a sequence of nonnegative weights whose sum is one, $\sum_{k=1}^n B_{nk}(\cdot) = 1$. The Nadaraya-Watson type of weights for $B_{nk}(\beta^T D_i)$ is used,

$$B_{nk}(\beta^T D_i) = \frac{K(\frac{\beta^T D_i - \beta^T D_k}{h_n})}{\sum_{l=1}^n K(\frac{\beta^T D_i - \beta^T D_l}{h_n})},$$

where h_n is the bandwidth such that $h_n \to 0$ as $n \to \infty$ and $K(\cdot)$ is a symmetric kernel function.

Step 3. Apply the least squares to the imputed log-transformed survival times for getting an updated estimator

$$\gamma_{a+1} = \left\{ \sum_{i=1}^n (D_i - \bar{D}_n)^{\otimes 2} \right\}^{-1} \sum_{i=1}^n (D_i - \bar{D}_n)\{\hat{Y}_i(\gamma_a) - \bar{Y}_n(\gamma_a)\},$$

where $\bar{Y}_n(\gamma_a) = n^{-1} \sum_{i=1}^n \hat{Y}_i(\gamma_a)$.

Step 4. Repeat Steps 2 and 3 until a certain convergence criteria is achieved. We denote by $\hat{\gamma}_B$ the converged estimator.

The local Buckely-James estimator $\hat{\gamma}_B$ is the solution to

$$\begin{aligned} U_{1,B}(\gamma) &= \sum_{i=1}^{n}(D_i - \bar{D}_n)\{\gamma^T D_i - \hat{Y}_i(\gamma)\} \\ &= \sum_{i=1}^{n}\left\{\int_{-\infty}^{\infty} t\, dY_i^D(t,\gamma) + \int_{-\infty}^{\infty}\int_{t}^{\infty} \frac{1-\hat{F}_{i\breve{\theta}}(s)}{1-\hat{F}_{i\breve{\theta}}(t)} ds dJ_i^D(t,\gamma)\right\} = 0, \end{aligned}$$

where $Y_i^D(t,\gamma) = (D_i - \bar{D}_n)1\{e_i(\gamma) \geq t\}$ and $J_i^D(t,\gamma) = (D_i - \bar{D}_n)1\{e_i(\gamma) \geq t, \Delta_i^Y = 0\}$, and $\hat{F}_{i\breve{\theta}}(t)$ is the shorthand notation of $\hat{F}_{\breve{\theta}}(t \mid \hat{\beta}_P^T D_i)$, where $\breve{\theta}^T = (\gamma^T, \hat{\beta}_P^T)$. Since $U_{1,B}(\gamma)$ is neither continuous nor monotone in γ, we define $\hat{\gamma}_B$ as a zero-crossing of $U_{1,B}(\gamma)$. Similar as in Lai and Ying (1991), we define $V_{1,B}(\gamma)$ as a smooth approximated function of $U_{1,B}(\gamma)$,

$$V_{1,B}(\gamma) = \sum_{i=1}^{n}\left\{\int_{-\infty}^{\infty} t\, dEY_i^D(t,\gamma) + \int_{-\infty}^{\infty}\int_{t}^{\infty} \frac{1-F_{i\tilde{\theta}_0}(s)}{1-F_{i\tilde{\theta}_0}(t)} ds dEJ_i^D(t,\gamma)\right\},$$

where $F_{i\tilde{\theta}_0}(t)$ is the limit of $\hat{F}_{i\breve{\theta}}$ and $\tilde{\theta}_0^T = (\gamma^T, \beta_0^T)$.

Pang et al. (2014) took the resampling technique of Jin et al. (2006b) to make an inference on $\hat{\gamma}_B$. The variance estimation using the resampling is very similar to that used in rank estimation. The key step of resampling method is to generate positive random variables R_i, which are independent of the data, $i = 1, ..., n$, with $E(R_i) = \text{var}(R_i) = 1$. We define

$$L^*(\gamma) = \left\{\sum_{i=1}^{n} R_i(D_i - \bar{D}_n)^{\otimes 2}\right\}^{-1}\left[\sum_{i=1}^{n} R_i(D_i - \bar{D}_n)\{\hat{Y}_i^*(\gamma) - \bar{Y}_n^*(\gamma)\}\right], \tag{2.20}$$

where

$$\begin{aligned} \hat{Y}_i^*(\gamma) &= \Delta_i^Y \tilde{Y}_i + (1-\Delta_i^Y)\left[\frac{\int_{e_i(\gamma)}^{\infty} u\, d\hat{F}_{\breve{\theta}}^*(u \mid \hat{\beta}_P^T D_i)}{1-\hat{F}_{\breve{\theta}}^*\{e_i(\gamma) \mid \hat{\beta}_P^T D_i\}} + \gamma^T D_i\right], \\ \hat{F}_{\breve{\theta}}^*(t \mid \hat{\beta}_P^T D_i) &= 1 - \prod_{j: e_j(\gamma) < t}^{n}\left\{1 - \frac{R_j B_{nj}(\hat{\beta}_P^T D_i)\Delta_j^Y}{\sum_{k=1}^{n} R_k I\{e_k(\gamma) \geq e_j(\gamma)\} B_{nk}(\hat{\beta}_P^T D_i)}\right\}, \end{aligned}$$

and $\bar{Y}_n^*(\gamma) = n^{-1}\sum_{i=1}^{n}\hat{Y}_i^*(\gamma)$. Note that $L^*(\gamma)$ is a perturbed version of local Bukley-James estimator. The updated perturbed estimator is obtained by

$$\gamma_{a+1}^{*} = L^{*}(\gamma_{a}^{*}).$$

We denote by $\hat{\gamma}_B^*$ the converged perturbed estimator.

To approximate the asymptotic distribution of $\hat{\gamma}_B$, we repeat the procedure described above many times to generate many of $\hat{\gamma}_B^*$. That is, we start with the Buckely-James estimator with an initial value and compute $L^*(\gamma)$ with being perturbed by $R_i(i = 1, ..., n)$. Repeating this procedure gives many of $\hat{\gamma}_B^*$. By Jin et al. (2006b) and Pang et al. (2014), the conditional distribution of $n^{1/2}(\hat{\gamma}_B^* - \hat{\gamma}_B)$ converges almost surely to the asymptotic distribution of $n^{1/2}(\hat{\gamma}_B - \gamma_0)$. Based on this result, the variance of $\hat{\gamma}_B$ can be estimated by the sample variance of $\hat{\gamma}_B^*$.

The parameter β_0 is estimated by the least squares estimator, namely,

$$\hat{\beta}_P = (D_{(c)}^T D_{(c)})^{-1} D_{(c)}^T Z.$$

The estimator $\hat{\beta}_P$ is the solution to

$$U_{2,P}(\beta) = \sum_{i=1}^{n} (D_i - \bar{D})(Z_i - D_i^T \beta).$$

Note that the reduced exposure model with the binary exposure has a heteroscedastic error variance. We used theoretical results of White (1980) to state the asymptotic properties of $\hat{\beta}_P$.

We can generate the resampled least squares by solving the following perturbed estimating equation,

$$U_{2,P}(\beta) = \sum_{i=1}^{n} (D_i - \bar{D})(Z_i - D_i^T \beta) R_i,$$

where $R_i(i = 1, ..., n)$ are the same random variables used in (2.20).

Corollary 4. *For Case 2-2, the local Buckley-James estimator for γ_0, denoted as $\hat{\gamma}_B$, and the least* *squares estimator for β_0, denoted as $\hat{\beta}_P$, satisfy Assumption 1 and 2 under the conditions B1-B9 in the appendix. Therefore the two-stage estimator, $\hat{\alpha}$, with $\hat{\gamma}_B$ and $\hat{\beta}_P$ converges in probability to α_0 and asymptotically normal by Theorem 1 and 2.*

Theorem 6. *For Case 2-2, under the conditions B1-B9 in the appendix, the asymptotic distribution of $\hat{\alpha}$ can be estimated by the empirical distribution of $\hat{\alpha}^* = (\hat{\alpha}^1, ..., \hat{\alpha}^K)$, conditionally on the data.*

2.4 Simulation study

We conducted extensive simulation studies to evaluate the proposed two-stage estimation in finite sample sizes. Four scenarios are considered to generate Case 1-1 to Case 2-2. For Case 1-1, Case 1-2 and Case 2-1, the simulation models are given by

$$Y_i = \alpha_{yo} + \alpha_{yz} Z_i + \alpha_{yw}^T W_i + \alpha_{yu} U_i + \varepsilon_i,$$
$$Z_i = \beta_{zo} + \beta_{zv}^T V_i + \beta_{zw}^T W_i + \beta_{zu} U_i + \delta_i.$$

We generated $(V_i^T, W_i^T, U_i)^T$ from a standard normal truncated at ± 2. Censoring time for Y_i were generated form Unif$(0, c_y)$, where c_y is determined to yield a desired right-censoring rate, 20%. For Case 1-1 and Case 1-2, we considered two-dimensional V_i and W_i, while for Case 2-1, we considered univariate V_i and W_i. All parameters are set to be one except for Case 2-1 where $\beta_{zo} = 0$ and the rest of parameters are one.

For Case 1-1, we generated ε_i and δ_i independently from standard normal, $N(0,1)$ and standard Gumbel, $G(0,1)$. The variable from $G(0,1)$ was standardized to have mean 0 and variance 1. Since Gumbel distribution has a skewness, which is about 1.14, we can evaluate the effect of skewness of error distributions on estimation. For Case 1-2, $\varepsilon_i \sim N(0,1)$ and $\delta_i \sim N(0,1)$ independently. Left-censored exposure was considered and the censoring variable was generated from Unif$(-7, c_z)$, where c_z is determined to yield a desired left-censoring rate, 20%. For Case 2-1, $U_i \sim N(0, 0.5)$, $\varepsilon_i \sim N(0,1)$ and $\delta_i \sim N(0, 0.5)$ so $\beta_{zu} U_i + \delta_i \sim N(0,1)$.

For Case 2-2, the same response model was used with a standard normal ε_i and univariate V_i ~ Beroulli(0.5), W_i ~ Beroulli(0.5) and U_i ~ Beroulli(0.5). All parameters were set to be one. The binary exposure Z_i was generated by Z_i ~ Beroulli(P_i) with

$$P_i = \beta_{zo} + \beta_{zv}^T V_i + \beta_{zw}^T W_i + \beta_{zu} U_i,$$

where $(\beta_{zo}, \beta_{zv}, \beta_{zw}, \beta_{zu}) = (0, b, 0.1, 0.2)$ and $b \in \{0.2, 0.4\}$.

The size of resampling to estimate the covariance matrix is 500. The perturbing variables were generated from an exponential distribution with mean 1. We estimated the parameters using the proposed two-stage estimation with the rank and the local Buckley-James estimators with optimal and identity weight matrices, and the naive rank-based method based on fitting the accelerated failure time model directly to the exposure and observed confounders. We used the R package called `lss` (Huang and Jin 2006) for implementation. We iterated this procedure 500 times. The results are summarized in Table 2.1-2.4. For Case 2-1 and Case 2-2, since B is invertible, the two-stage estimator does not depend on A_n.

The simulation results show that the proposed estimator is unbiased and the variance estimator obtained by the proposed resampling method performs well. The IV estimators with the identity matrix and optimal weights performed similarly. The naive method gave biased estimators and the confidence intervals for the exposure effect were far below the target coverage rate, 95%.

In Case 1-1 results, the naive estimators have much greater biases when the exposure is generated from Gumbel distribution. This implies that skewness of exposure distribution enlarges the bias of the naive estimators. The two-stage estimators performed well across a range of the response and exposure distributions we considered. In Case 2-1, we conducted complete case naive analysis. There is a bias in that analysis because exogenous censoring does not hold as an unmeasured confounder exists. Thus, the method of Wang and Feng (2012) also will not be reliable in our simulation setting. In Case 2-2 results with $b = 0.2$, the proposed estimator based on the local Buckely-James estimation is unbiased, however, the estimator based on rank estimation has a bias, and that bias decreases as b increases to 0.4.

2.5 SEER Colon Cancer Data

We applied the proposed method to Surveillance, Epidemiology and End Results (SEER) data for elderly stage III colon cancer patients. Oxaliplatin is a chemotherapeutic agent that is used as part of a multi-agent adjuvant chemotherapy regimen for stage III colon cancer patients. Based on 2003 efficacy results from the MOSAIC trial (Andre et al. 2004), the FDA approved Oxaliplatin for use in stage III colon cancer. After FDA approval for this new indication, it disseminated rapidly among stage III colon cancer patients to replace 5-fluorouracil (FU) monotherapy as the standard of care. The objective of our analysis is to determine if Oxaliplatin, compared to 5-FU alone, will increase survival time of the patients.

The cohort included individuals aged 65+ from 12 US states who were diagnosed with primary stage III colon cancer between 2003 and 2007, with follow-up through April 2010. Included patients received surgical resection within 90 days of diagnosis, survived longer than 30 days, and initiated either Oxaliplatin or 5-FU/capecitabine without Oxaliplatin within 110 days of surgery and 120 days of diagnosis. Patients who received radiation, were diagnosed at autopsy, or had HMO coverage or incomplete Medicare claims during the 12 months pre- and post-diagnosis (or until death) were excluded.

The binary exposure variable (Trt) is coded as 1 if the patient was treated with Oxaliplatin and 0 if being treated with 5-FU. The IV (Time) is coded as 1 if the patient was treated after the US Food and Drug Administration's approval of Oxaliplatin for use in Stage III colon cancer (Mack et al. 2015) and 0 otherwise. The three confounders were used: age in years (age), household median income in 2000 in 10,000 dollar units (income) and an indicator for diabetes (Dia). We generated dummy variables for age and income as follows. Four categories for age were generated using quartiles of age distribution: Group.1 = $I(65 \leq \text{age} < 70)$, Group.2 = $I(70 \leq \text{age} < 74)$, Group.3 = $I(74 \leq \text{age} < 78)$ and Group.4 = $I(78 \leq \text{age})$. And then, three dummy variables to compare Group.j to Group.1 were generated: Age.j = $I(\text{age} \in \text{Group}.j)$

Table 2.1: Results for Case 1-1 (NN): empirical bias (Bias), empirical standard error (ESE), average of the estimated standard error (ASE) and empirical coverage rate (ECR) of 95% Wald-type confidence interval.

n	Methods	Parameter	Bias	ESE	ASE	ECR
100	Identity matrix	α_{yz}	-0.002	0.140	0.147	0.956
		$\alpha_{yw,1}$	-0.001	0.231	0.254	0.950
		$\alpha_{yw,2}$	0.017	0.243	0.252	0.962
	Optimal weight	α_{yz}	-0.004	0.142	0.145	0.960
		$\alpha_{yw,1}$	-0.002	0.235	0.249	0.944
		$\alpha_{yw,2}$	0.015	0.247	0.248	0.948
	Naive	α_{yz}	0.237	0.081	0.079	0.172
		$\alpha_{yw,1}$	-0.246	0.177	0.180	0.706
		$\alpha_{yw,2}$	-0.229	0.183	0.178	0.712
200	Identity matrix	α_{yz}	-0.002	0.097	0.099	0.954
		$\alpha_{yw,1}$	0.008	0.164	0.170	0.950
		$\alpha_{yw,2}$	0.007	0.159	0.170	0.956
	Optimal weight	α_{yz}	-0.003	0.098	0.098	0.954
		$\alpha_{yw,1}$	0.008	0.165	0.168	0.950
		$\alpha_{yw,2}$	0.008	0.160	0.168	0.952
	Naive	α_{yz}	0.232	0.056	0.055	0.018
		$\alpha_{yw,1}$	-0.230	0.127	0.125	0.542
		$\alpha_{yw,2}$	-0.228	0.121	0.125	0.574
400	Identity weight	α_{yz}	0.004	0.068	0.069	0.952
		$\alpha_{yw,1}$	-0.002	0.111	0.118	0.972
		$\alpha_{yw,2}$	-0.003	0.119	0.117	0.926
	Optimal weight	α_{yz}	0.003	0.068	0.069	0.952
		$\alpha_{yw,1}$	-0.002	0.112	0.118	0.968
		$\alpha_{yw,2}$	-0.002	0.120	0.117	0.930
	Naive	α_{yz}	0.237	0.038	0.039	0.000
		$\alpha_{yw,1}$	-0.236	0.085	0.088	0.234
		$\alpha_{yw,2}$	-0.234	0.088	0.089	0.250

Note: “Identity matrix" is the proposed two-stage IV method with $A_n = I$. “Optimal weight" is the proposed two-stage IV method with $A_n = \hat{\Omega}(\hat{\alpha}_I)$. “Naive" is the rank method with observed variables.

Table 2.2: Results for Case 1-2: empirical bias (Bias), empirical standard error (ESE), average of the estimated standard error (ASE) and empirical coverage rate (ECR) of 95% Wald-type confidence interval.

n	Methods	Parameter	Bias	ESE	ASE	ECR
100	Identity matrix	α_{yz}	0.000	0.157	0.154	0.936
		$\alpha_{yw,1}$	0.012	0.255	0.255	0.936
		$\alpha_{yw,2}$	-0.017	0.271	0.255	0.936
	Optimal weight	α_{yz}	-0.001	0.158	0.152	0.937
		$\alpha_{yw,1}$	0.012	0.256	0.250	0.937
		$\alpha_{yw,2}$	-0.018	0.272	0.251	0.937
	Naive	α_{yz}	0.236	0.101	0.096	0.298
		$\alpha_{yw,1}$	-0.227	0.204	0.204	0.766
		$\alpha_{yw,2}$	-0.246	0.204	0.206	0.744
200	Identity matrix	α_{yz}	0.001	0.099	0.104	0.951
		$\alpha_{yw,1}$	-0.004	0.172	0.175	0.951
		$\alpha_{yw,2}$	-0.003	0.182	0.175	0.951
	Optimal weight	α_{yz}	0.001	0.099	0.103	0.945
		$\alpha_{yw,1}$	-0.005	0.171	0.173	0.945
		$\alpha_{yw,2}$	-0.004	0.182	0.173	0.945
	Naive	α_{yz}	0.233	0.063	0.066	0.046
		$\alpha_{yw,1}$	-0.241	0.134	0.142	0.616
		$\alpha_{yw,2}$	-0.232	0.155	0.142	0.610
400	Identity matrix	α_{yz}	0.001	0.076	0.073	0.947
		$\alpha_{yw,1}$	-0.002	0.121	0.122	0.947
		$\alpha_{yw,2}$	-0.002	0.124	0.123	0.947
	Optimal weight	α_{yz}	0.000	0.076	0.073	0.945
		$\alpha_{yw,1}$	-0.002	0.121	0.122	0.945
		$\alpha_{yw,2}$	-0.002	0.123	0.122	0.945
	Naive	α_{yz}	0.237	0.046	0.047	0.002
		$\alpha_{yw,1}$	-0.237	0.103	0.101	0.344
		$\alpha_{yw,2}$	-0.236	0.101	0.101	0.344

Note: “Identity matrix" is the proposed two-stage IV method with $A_n = I$. “Optimal weight" is the proposed two-stage IV method with $A_n = \hat{\Omega}(\hat{\alpha}_I)$. “Naive" is the rank method with observed variables.

Table 2.3: Results for Case 2-1: empirical bias (Bias), empirical standard error (ESE), average of the estimated standard error (ASE) and empirical coverage rate (ECR) of 95% Wald-type confidence interval.

n	Methods	Parameter	Bias	ESE	ASE	ECR
100	Two-stage	α_{yz}	-0.025	0.245	0.254	0.946
		α_{yw}	0.002	0.304	0.298	0.954
	Naive	α_{yz}	0.289	0.101	0.101	0.186
		α_{yw}	-0.288	0.181	0.176	0.622
200	Two-stage	α_{yz}	-0.025	0.177	0.174	0.917
		α_{yw}	0.006	0.208	0.203	0.933
	Naive	α_{yz}	0.285	0.068	0.071	0.012
		α_{yw}	-0.281	0.123	0.123	0.378
400	Two-stage	α_{yz}	-0.003	0.117	0.123	0.958
		α_{yw}	-0.005	0.137	0.142	0.970
	Naive	α_{yz}	0.279	0.047	0.050	0.000
		α_{yw}	-0.277	0.088	0.086	0.130

Note: “Two-stage" is the proposed two-stage IV method. “Naive" is the rank method with observed variables.

for $j = 2, ..., 4$. A dummy variable for income is generated by Inc = $I\{\text{income} \geq 50,000\}$. The response ($Y$) is log survival time in years. The sample size is 2879, with the resampling size equal to 200 when computing the standard errors of the parameter estimates with unit exponential perturbations.

The models to be estimated are given by

$$Y = \alpha^*_{yz}\text{Trt}^* + \alpha^*_{yw,1}\text{Age.2} + \alpha^*_{yw,2}\text{Age.3} + \alpha^*_{yw,3}\text{Age.4} + \alpha^*_{yw,4}\text{Inc} + \alpha^*_{yw,5}\text{Dia} + \text{e}^*_y, \qquad (2.21)$$

$$Y = \alpha_{yz}\text{Trt} + \alpha_{yw,1}\text{Age.2} + \alpha_{yw,2}\text{Age.3} + \alpha_{yw,3}\text{Age.4} + \alpha_{yw,4}\text{Inc} + \alpha_{yw,5}\text{Dia} + \text{e}_y, \qquad (2.22)$$

where Trt* in (2.21) is an unobserved latent variable for Trt, which is assumed to be observed via Trt = $I(\text{Trt}^* > 0)$. To account for possible endogeneity of Trt* and Trt, we conducted IV analyses of Case 2-1 and Case 2-2 with the rank, Buckley-James and local Buckely-James estimators.

The corresponding exposure models to (2.21) and (2.22) are given by

$$\text{Trt}^* = \beta^*_{zo} + \beta^*_{zv}\text{Time} + \beta^*_{zw,1}\text{Age.2} + \beta^*_{zw,2}\text{Age.3}+$$

Table 2.4: Results for Case 2-2: empirical bias (Bias), empirical standard error (ESE), average of the estimated standard error (ASE) and empirical coverage rate (ECR) of 95% Wald-type confidence interval.

			b = 0.2				b = 0.4			
n	Methods	Para	Bias	ESE	ASE	ECR	Bias	ESE	ASE	ECR
800	LBJ	α_{yz}	-0.035	0.257	0.252	0.962	-0.013	0.124	0.121	0.948
		α_{yw}	0.002	0.054	0.055	0.962	-0.002	0.048	0.049	0.952
	Rank	α_{yz}	-0.152	0.264	0.261	0.920	-0.024	0.143	0.138	0.926
		α_{yw}	0.003	0.056	0.056	0.966	-0.001	0.051	0.053	0.958
	Naive	α_{yz}	0.288	0.057	0.059	0.002	0.234	0.053	0.053	0.006
		α_{yw}	-0.027	0.044	0.046	0.906	-0.023	0.044	0.046	0.926
1600	LBJ	α_{yz}	-0.010	0.177	0.173	0.949	-0.004	0.087	0.085	0.936
		α_{yw}	0.001	0.038	0.038	0.947	0.000	0.034	0.035	0.959
	Rank	α_{yz}	-0.134	0.184	0.182	0.896	-0.014	0.098	0.097	0.942
		α_{yw}	0.002	0.038	0.039	0.952	0.000	0.036	0.037	0.964
	Naive	α_{yz}	0.292	0.039	0.042	0.000	0.233	0.037	0.037	0.000
		α_{yw}	-0.028	0.032	0.032	0.852	-0.021	0.031	0.032	0.902
3200	LBJ	α_{yz}	0.000	0.118	0.123	0.952	0.002	0.061	0.060	0.950
		α_{yw}	0.000	0.027	0.027	0.960	-0.001	0.025	0.025	0.958
	Rank	α_{yz}	-0.124	0.124	0.128	0.844	-0.009	0.069	0.069	0.957
		α_{yw}	0.000	0.027	0.028	0.954	-0.001	0.026	0.026	0.961
	Naive	α_{yz}	0.289	0.029	0.029	0.000	0.232	0.026	0.026	0.000
		α_{yw}	-0.027	0.023	0.023	0.786	-0.023	0.023	0.023	0.836

Note: “LBJ" is the proposed two-stage IV method with local Buckley-James estimator. “Rank" is the proposed two-stage IV method with rank estimator. “Naive" is the rank method with observed variables.

$$\beta^*_{zw,3}\text{Age.4} + \beta^*_{zw,5}\text{Inc} + \beta^*_{zw,6}\text{Dia} + \text{e}^*_z, \tag{2.23}$$

$$\text{Trt} = \beta_{zo} + \beta_{zv}\text{Time} + \beta_{zw,1}\text{Age.2} + \beta_{zw,2}\text{Age.3}+$$

$$\beta_{zw,3}\text{Age.4} + \beta_{zw,5}\text{Inc} + \beta_{zw,6}\text{Dia} + \text{e}_z, \tag{2.24}$$

where e^*_z is assumed to follow a standard normal distribution and e_z is assumed to follow a Bernoulli distribution with the mean zero and the variance being a function of the conditional mean of Trt.

The results are given in Table 2.5. Since there is only a single IV, B_0 is invertible and the two-stage estimators do not depend on A_n. The naive estimate of Trt is 0.149 with p-value of 0.059, hence was not significant at level of 0.05. In contrary, all of the two-stage estimates of Trt* and Trt were significant at level of 0.05. The parameter α_{yz} has an interpretation that Oxaliplatin, compared to 5-FU, increases median survival time by $100(e^{\alpha_{yz}} - 1)\%$ if α_{yz} is positive. The amounts of increase in median survival time were 44% (9%, 91%) from rank estimator, 48% (11%, 97%) from Buckley-James estimator, and 49% (12%, 98%) from local Buckley-James estimator. Based on the two-stage IV analysis, we conclude that Oxaliplatin is more beneficial than 5-FU in treating colon cancer patients. The difference in naive and IV analyses suggests that there may exist unmeasured confounders.

The F-statistic for Time variable in (2.24) is 1328.75, thus it is much greater than rule of thumb of 10 (Staiger and Stock 1997), which may imply that Time is a strong instrument. We examined whether the unmeasured confounding is similar between the IV groups by computing two parameters for the exposure effect, one for pre-FDA approval and the other for post-FDA approval. For this examination, we stratified the data by Time variable and calculated the naive estimate for each stratum. Similar inference results for those two estimates may indicate that the unmeasured confounder is balanced well between the two IV periods. The estimate of Trt for pre-FDA approval was 0.059 with p-value of 0.706 and that for post-FDA approval was 0.004 with p-value of 0.976. Thus, the inference results are similar and this supports the validity of our IV analysis.

The MOSAIC trial (Andre et al. 2004) provided the evidence for the improvement of the treatment of colon cancer by adding Oxaliplatin to a regimen of Fluorouracil and Leucovorin (FL). The primary outcome was disease-free survival. The study showed that the probabilities of disease-free survival at 3 years are 78.2% (75.6% - 80.7%) for the Oxaliplatin group and 72.9% (70.2% - 75.7%) for the FL group. The study also investigated the overall survival of stage III colon cancer patients in terms of a hazard ratio, which was 0.86 with a confidence interval of (0.66 - 1.11). Thus it was not significant at level of 0.05, while our result in the SEER data was significant.

2.6 Discussion

The proposed methods can be applied to other types of coarsened data if reliable estimation methods to fit the reduced models exist. For example, we may apply our method to the current status or general interval-censored data. Several methods have been developed for the accelerated failure time model with currents status or interval censored data; see for example Self and Grossman (1986), Rabinowitz et al. (1995), Betensky et al. (2011), and Tian and Cai (2006). Among them, the approaches of Betensky et al. (2011) and Tian and Cai (2006) can be readily adapted in our method.

We developed the method for Case 2-1, where both the response and exposure are censored. We illustrated out method in the context of left-censored exposure due to a limit of detection, but the method is not limited to left-censoring. One of interesting examples to have right-censored exposure is an observational duration-response study, where duration may be right-censored (Johnson and Tsiatis 2004). In may cases, treatment duration is left to the discretion of the physicians or investigators but often terminated by treatment-terminating events. Suppose we are interested in the effect of assigned treatment duration on the outcome. In this case, the censoring time for the treatment duration is the time to a treatment-terminating event and that for the outcome is the time to endpoint at that censoring time. Keeping the conditional independence assumption of censoring times and failure times made in Case 2-1 method, the proposed method

Table 2.5: Estimates (Est), Standard errors (Se), p-values (p-value) and 95% Wald confidence intervals (95% CI) for the parameters in the SEER data.

Methods	Parameter	Est	Se	p-value	95 % CI
Rank	Trt*	0.115	0.045	0.010	0.028 to 0.203
	Age.2	-0.138	0.102	0.176	-0.339 to 0.062
	Age.3	-0.264	0.115	0.021	-0.489 to -0.040
	Age.4	-0.631	0.111	0.000	-0.849 to -0.414
	Inc	0.148	0.072	0.040	0.006 to 0.289
	Dia	-0.208	0.087	0.016	-0.378 to -0.038
Rank	Trt	0.368	0.142	0.010	0.089 to 0.647
	Age.2	-0.140	0.102	0.171	-0.340 to 0.060
	Age.3	-0.270	0.115	0.018	-0.495 to -0.045
	Age.4	-0.650	0.108	0.000	-0.863 to -0.438
	Inc	0.151	0.072	0.035	0.011 to 0.292
	Dia	-0.210	0.087	0.015	-0.379 to -0.040
BJ	Trt	0.392	0.147	0.008	0.105 to 0.679
	Age.2	-0.122	0.106	0.250	-0.329 to 0.086
	Age.3	-0.310	0.121	0.010	-0.547 to -0.073
	Age.4	-0.656	0.117	0.000	-0.886 to -0.426
	Inc	0.152	0.076	0.045	0.003 to 0.300
	Dia	-0.245	0.093	0.009	-0.427 to -0.062
LBJ	Trt	0.398	0.144	0.006	0.116 to 0.681
	Age.2	-0.115	0.102	0.258	-0.314 to 0.084
	Age.3	-0.312	0.121	0.010	-0.549 to -0.075
	Age.4	-0.637	0.117	0.000	-0.866 to -0.408
	Inc	0.158	0.075	0.034	0.012 to 0.304
	Dia	-0.244	0.093	0.009	-0.425 to -0.062
Naive	Trt	0.149	0.079	0.059	-0.006 to 0.304
	Age.2	-0.162	0.102	0.112	-0.362 to 0.038
	Age.3	-0.294	0.114	0.010	-0.517 to -0.071
	Age.4	-0.709	0.104	0.000	-0.913 to -0.505
	Inc	0.167	0.070	0.017	0.030 to 0.304
	Dia	-0.220	0.085	0.010	-0.387 to -0.053

Note: “BJ" is the proposed two-stage IV method with Buckley-James estimator. “LBJ" is the proposed two-stage IV method with local Buckley-James estimator. “Rank" is the proposed two-stage IV method with rank estimator. “Naive" is the rank method with observed variables.

can be applied to estimate the parameters of interest. Potential IVs would be 'preference-based instruments' such as geographic region, hospital, dialysis center or individual physician (Brookhart et al. 2010).

Minimum distance estimators for causal parameters are derived from the reduced linear models. If appropriate estimators of the reduced linear models are available for any type of coarsened data, then causal inference will be easily made by using the proposed method. For the binary exposure (Case 2-2), we used the linear probability model. One is tempted to use other models such as a logistic regression model for the binary exposure and this approach sets both the reduced models to be non-linear. For a such non-linear case, it is not straightforward to derive minimum distance estimators without strong model assumptions. To obtain IV estimators in this case, we may use two-stage predictor substitution method (Terza et al. 2008), where the binary exposure is replaced by its predicted value of the exposure. However, in order to state its asymptotic properties, we need to identity other types of assumptions and theories. Further research for the extension of our method to the non-linear case would be needed.

2.7 Proofs

Before the proofs, we present some definitions for the rank estimators with right censored data. For $k = 1, 2$, let

$$M_i(\theta_k; t) = N_i(\theta_k; t) - \int_0^t Y_i(\theta_k; u)\lambda_k(u)du,$$

where $\theta^T = (\theta_1^T, \theta_2^T) = (\gamma^T, \beta^T)$, λ_1 and λ_2 are the common hazard functions of τ_i and δ_i^* respectively. The equation $\sum_{i=1}^n Y_i(\theta_k; t)\{D_i - \bar{D}(\theta_k; t)\} = 0$ implies that

$$U_{k,G}(\theta_k) = \sum_{i=1}^n \int_{-\infty}^{\infty} S^{(0)}(\theta_k; t)\{D_i - \bar{D}(\theta_k; t)\}dM_i(\theta_k; t),$$

where $\bar{D}(\theta_k;t) = S^{(1)}(\theta_k;t)/S^{(0)}(\theta_k;t)$. It is well known that $E\{M_i(\theta_{0k};t)\} = 0$, where $\theta_0^T = (\theta_{01}^T, \theta_{02}^T) = (\gamma_0^T, \beta_0^T)$ is the vector of true value of θ. Define

$$A_k = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{\infty} s^{(0)}(\theta_{0k};t)\{D_i - \bar{d}(\theta_{0k};t)\}^{\otimes 2}\{\dot{\lambda}_k(t)/\lambda_k(t)\}dN_i(\theta_{0k};t),$$

where $a^{\otimes 2} = aa^T$, $\dot{\lambda}(t) = d\lambda(t)/dt$, $s^{(0)}(\theta_{0k};t) = \lim_{n\to\infty} S^{(0)}(\theta_{0k};t)$ and $s^{(1)}(\theta_{0k};t) = \lim_{n\to\infty} S^{(1)}(\theta_{0k};t)$, $\bar{d}(\theta_{0k};t) = s^{(1)}(\theta_{0k};t)/s^{(0)}(\theta_{0k};t)$ for $k = 1,2$, and

$$\Sigma_{k,G} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} u_{k,G(i)}(\theta_{0k}) u_{k,G(i)}(\theta_{0k})^T,$$

where

$$u_{k,G(i)}(\theta_{0k}) = \int_{-\infty}^{\infty} s^{(0)}(\theta_{0k})\{D_i - \bar{d}(\theta_{0k};t)\}dM_i(\theta_{0k};t).$$

We impose the following regularity conditions.

Condition 1. *For $k = 1,2$, the parameter space Θ_k containing θ_k is compact.*

Condition 2. *For $i = 1,...,n$, the Euclidean norm $\|D_i\|$ is bounded by a constant.*

Condition 3. *Let $f_k(t)$ be the density function associated with $\lambda_k(t)(k = 1,2)$. Then $f_k(t)$ and $df_k(t)/dt$ are bounded and $\int(dlogf_k(u)/du)^2 f_k(u)du < \infty$.*

Condition 4. *The matrix $A_k(k = 1,2)$ is non-singular.*

Remark 4. *Condition 1–4 are the regularity conditions for the consistency and asymptotic normality of the Gehan rank estimators hold (Jin et al. 2006a, Ying 1993).*

Proof Corollary 1: Under Condition 1-4, by proof of theorem 1 of Jin et al. (2006a), the Gehan rank estimator, $\hat{\gamma}_G$, converges to γ_0 almost surely. The least squares estimator, $\hat{\beta}_L$ also converges to β_0 almost surely. Thus Assumption 1 holds.

Under condition 1-4, by using the arguments of Ying (1993), we have

$$n^{\frac{1}{2}}(\hat{\gamma}_G - \gamma_0) = -A_1^{-1} n^{-\frac{1}{2}} U_{1,G}(\gamma_0) + o_p(1). \tag{2.25}$$

A Taylor expansion of $U_{2,L}(\hat{\beta}_L)$ at β_0 gives

$$n^{\frac{1}{2}}(\hat{\beta}_L - \beta_0) = \bar{M}^{-1} n^{-\frac{1}{2}} U_{2,L}(\beta_0) + o_p(1). \tag{2.26}$$

The asymptotic covariance matrix of $n^{-\frac{1}{2}} U_{2,L}(\beta_0)$ is $\Sigma_{2,L} = \bar{M}\mathrm{var}(\delta^*)$.

By Lemma 1 of Jin et al. (2006a),

$$n^{-\frac{1}{2}} U_{1,G}(\gamma_0) = n^{-\frac{1}{2}} u_{1,G}(\gamma_0) + o_p(1). \tag{2.27}$$

By the uniform strong law of large numbers (Pollard 1990), $s^{(0)}(\gamma_0; t) = \lim_{n\to\infty} S^{(0)}(\gamma_0; t)$ and $s^{(1)}(\gamma_0; t) = \lim_{n\to\infty} S^{(1)}(\gamma_0; t)$. The asymptotic covariance matrix of $u_{1,G}(\gamma_0)$ is $\Sigma_{1,G}$.

The asymptotic covariance matrix of $n^{-\frac{1}{2}} u_{1,G}(\gamma_0)$ and $n^{-\frac{1}{2}} U_{2,L}(\beta_0)$ is given by

$$\Sigma_{G,L} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} u_{1,G(i)}(\gamma_0) U_{2,L(i)}(\beta_0)^T, \tag{2.28}$$

where $U_{2,L(i)}(\beta_0) = (Z_i - D_i^T \beta_0)(D_i - \bar{D})$.

In view of (2.25), (2.26), (2.27) and (2.28), the multivariate central limit theorem implies that $n^{1/2}\{(\hat{\gamma}_G, \hat{\beta}_L) - (\gamma_0, \beta_0)\}$ converges to a mean-zero normal distribution with the covariance matrix

$$\Sigma_{GL} = \begin{pmatrix} A_1^{-1}\Sigma_{1,G}A_1^{-1} & A_1^{-1}\Sigma_{G,L}\bar{M}^{-1} \\ \bar{M}^{-1}\Sigma_{G,L}^T A_1^{-1} & \bar{M}^{-1}\Sigma_{2,L}\bar{M}^{-1} \end{pmatrix}.$$

Proof Corollary 2: In order to get the two-stage estimator, we fit the two reduced models. We can build the multivariate model to represent those reduced models,

$$H_{ik} = \theta_{0k}^T D_i + \eta_{ik}, \; k = 1, 2,$$

where $(H_{i1}, H_{i2}) = (Y_i, Z_i)$, $(\theta_{01}^T, \theta_{02}^T) = (\gamma_0^T, \beta_0^T)$ and $(\eta_{i1}, \eta_{i2}) = (\tau_i, \delta_i^*)$. If both response and exposure are censored, then this model is a multivariate AFT model with multiple event data where the design matrices are identical for all types of the events (Jin et al. 2006a). In this case, we have two types of events ($k = 1, 2$). We fit this multivariate AFT model with the methods of Jin et al. (2006a), which are used for Case 1-2. Then, by Theorem 1 of Jin et al. (2006a), Assumption 1 and 2 hold.

Proof Corollary 3: Under the regularity conditions for the MLE, the MLE for β_0, $\hat{\beta}_M$, converges to β_0 almost surely. We also have asymptotic linearity to state,

$$\begin{aligned} n^{\frac{1}{2}}(\hat{\beta}_M - \beta_0) &= [0_{(p+q)\times 1}\ I_{p+q}] I(\beta_{zo}, \beta_0)^{-1} n^{-\frac{1}{2}} U_{2,M}(\beta_{zo}, \beta_0) + o_p(1), \\ &= I^*(\beta_{zo}, \beta_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} U_{2,M(i)}(\beta_{zo}, \beta_0) + o_p(1), \end{aligned} \tag{2.29}$$

where $U_{2,M(i)}(\beta_{zo}, \beta_0)$ is the ith contribution to the probit score function, $I(\beta_{zo}, \beta_0)$ is the corresponding Fisher information matrix and $I^*(\beta_{zo}, \beta_0)^{-1} = [0_{(p+q)\times 1}\ I_{p+q}] I(\beta_{zo}, \beta_0)^{-1}$. The function $I(\beta_{zo}, \beta_0)$ is defined as

$$I(\beta_{zo}, \beta_0) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} U_{2,M(i)}(\beta_{zo}, \beta_0) U_{2,M(i)}(\beta_{zo}, \beta_0)^T.$$

The asymptotic covariance matrix of $n^{-\frac{1}{2}} U_{2,M}(\beta_{zo}, \beta_0)$ is $I(\beta_{zo}, \beta_0)$. The asymptotic covariance matrix of $n^{-\frac{1}{2}} u_{1,G}(\gamma_0)$ and $n^{-\frac{1}{2}} U_{2,M}(\beta_0)$ is given by

$$\Sigma_{G,M} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} u_{1,G(i)}(\gamma_0) U_{2,M(i)}(\beta_{zo}, \beta_0)^T. \tag{2.30}$$

In view of (2.25), (2.27), (2.29) and (2.30), the multivariate central limit theorem implies that $n^{1/2}\{(\hat{\gamma}_G, \hat{\beta}_M) - (\gamma_0, \beta_0)\}$ converges to mean-zero normal distribution with the covariance matrix

$$\Sigma_{GL} = \begin{pmatrix} A_1^{-1} \Sigma_{1,G} A_1^{-1} & A_1^{-1} \Sigma_{G,M} \mathrm{tr}\{I^*(\beta_{zo}, \beta_0)^{-1}\} \\ I^*(\beta_{zo}, \beta_0)^{-1} \Sigma_{G,M}^T A_1^{-1} & I(\beta_0)^{-1} \end{pmatrix},$$

where $I(\beta_0)^{-1} = [0_{(p+q)\times 1}\ I_{p+q}] I(\beta_{zo}, \beta_0)^{-1} [0_{(p+q)\times 1}\ I_{p+q}]^T$.

Proof Theorem 3: To prove Theorem 3, it suffices to show that the conditional distribution of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ converges to the limiting distribution of $n^{1/2}(\hat{\theta} - \theta_0)$, where $\hat{\theta}^T = (\hat{\gamma}_G^T, \hat{\beta}_L^T)$ and $\hat{\theta}^{*T} = (\hat{\gamma}_G^{*T}, \hat{\beta}_L^{*T})$ is the solution of the perturbed estimating equations.

The loss functions for reduced response and exposure models are given by

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i^Y\{e_i(\gamma) \le e_j(\gamma)\}^-, \; \sum_{i=1}^{n}(Z_i - \beta_o - D_i^T\beta)^2,$$

where β_o is an intercept, $a^- = I(a < 0) \mid a \mid$ and the two loss functions are convex. Suppose we perturb the two estimation equations with the same positive random variables $R_i (i = 1, ..., n)$ whose mean and variance are all one. Then perturbed loss functions are given by

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i^Y\{e_i(\gamma) \le e_j(\gamma)\}^- R_i R_j, \; \sum_{i=1}^{n}(Z_i - \beta_o - D_i^T\beta)^2 R_i.$$

By the way of random perturbation, the perturbed loss functions retain the convexities of the original loss functions. Thus, the consistency of $\hat{\theta}^*$ can be proven by the same arguments to prove the consistency of $\hat{\theta}$ (Jin et al. 2006a).

Jin et al. (2006a) showed that the equation (10) can be written as

$$U_{1,G}^*(\gamma) = \sum_{i=1}^{n}\int_{-\infty}^{\infty} S^{(0)}(\gamma;t)\{D_i - \bar{D}(\gamma;t)\}dM_i(\gamma;t)R_i. \tag{2.31}$$

This is the critical result to derive the joint distribution of $\hat{\theta}^*$. From the results of Jin et al. (2006a), it follows that

$$n^{\frac{1}{2}}(\hat{\gamma}_G^* - \hat{\gamma}_G) = -A_1^{-1}n^{-\frac{1}{2}}U_{1,G}^*(\hat{\gamma}_G) + o_p(1), \tag{2.32}$$

$$n^{\frac{1}{2}}(\hat{\beta}_L^* - \hat{\beta}_L) = \bar{M}^{-1}n^{-\frac{1}{2}}U_{2,L}^*(\hat{\beta}_L) + o_p(1), \tag{2.33}$$

and

$$U_{1,G}^*(\hat{\gamma}_G) = U_{1,G}^*(\hat{\gamma}_G) - U_{1,G}(\hat{\gamma}_G) + o(n^{1/2}),\ U_{2,L}^*(\hat{\beta}_L) = U_{2,L}^*(\hat{\beta}_L) - U_{2,L}(\hat{\beta}_L) + o(n^{1/2}), \tag{2.34}$$

as $\hat{\gamma}_G$ and $\hat{\beta}_L$ are roots of $U_{1,G}(\gamma)$ and $U_{2,L}(\beta)$. From (2.31) and (2.34),

$$n^{-\frac{1}{2}}U_{1,G}^*(\hat{\gamma}_G) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_{-\infty}^{\infty} S^{(0)}(\hat{\gamma}_G;t)\{D_i - \bar{D}(\hat{\gamma}_G;t)\}dM_i(\hat{\gamma}_G;t)(R_i-1) + o(1), \tag{2.35}$$

$$n^{-\frac{1}{2}}U_{2,L}^*(\hat{\beta}_L) = n^{-\frac{1}{2}}\sum_{i=1}^{n}(Z_i - D_i^T\hat{\beta}_L)(D_i - \bar{D})(R_i-1) + o(1). \tag{2.36}$$

Conditionally on the data, the right-hand sides of (2.35) and (2.36) are the sum of independent zero-mean random vectors. Therefore, the multivariate central limit theorem implies that the conditional distribution of $n^{-1/2}\{U_{1,G}^*(\hat{\gamma}_G)^T, U_{2,L}^*(\hat{\beta}_L)^T\}^T$ given the data converges to a mean 0 multivariate normal distribution with the covariance matrix

$$\tilde{\Sigma}_{GL} = \begin{pmatrix} \tilde{\Sigma}_{1,G} & \tilde{\Sigma}_{G,L} \\ \tilde{\Sigma}_{G,L}^T & \tilde{\Sigma}_{2,L} \end{pmatrix},$$

where

$$\tilde{\Sigma}_{1,G} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} U_{1,G(i)}(\hat{\gamma}_G)U_{1,G(i)}(\hat{\gamma}_G)^T,$$
$$\tilde{\Sigma}_{2,L} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} U_{2,L(i)}(\hat{\beta}_L)U_{2,L(i)}(\hat{\beta}_L)^T,$$
$$\tilde{\Sigma}_{G,L} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} U_{1,G(i)}(\hat{\gamma}_G)U_{2,L(i)}(\hat{\beta}_L)^T,$$

and

$$U_{1,G(i)}(\hat{\gamma}_G) = \int_{-\infty}^{\infty} S^{(0)}(\hat{\gamma}_G;t)\{D_i - \bar{D}(\hat{\gamma}_G;t)\}dM_i(\hat{\gamma}_G;t),$$
$$U_{2,L(i)}(\hat{\beta}_L) = (Z_i - D_i^T\hat{\beta}_L)(D_i - \bar{D}).$$

As $S^{(0)}(\gamma;t) \to s^{(0)}(\gamma;t)$, $S^{(1)}(\gamma;t) \to s^{(1)}(\gamma;t)$ and $\hat{\theta} \to \theta_0$, we have $\tilde{\Sigma}_{GL} \to \Sigma_{GL}$. Then it follows from (2.32) and (2.33) that the conditional distribution of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ converges to a mean 0 multivariate normal distribution with the covariance matrix Σ_{GL}, which is the limiting distribution of $n^{1/2}(\hat{\theta} - \theta_0)$.

Proof Theorem 4: As we mentioned in proof of Corollary 2, for Case 1-2, the model we fit is a multivariate AFT model with multiple event data. We fit this model using the method of Jin et al. (2006a) which is based on marginal linear models. By Theorem 1 of Jin et al. (2006a), the conditional distribution of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ converges to the limiting distribution of $n^{1/2}(\hat{\theta} - \theta_0)$, where $\hat{\theta}^T = (\hat{\gamma}_G^T, \hat{\beta}_G^T)$ and $\hat{\theta}^{*T} = (\hat{\gamma}_G^{*T}, \hat{\beta}_G^{*T})$ is the solution of the perturbed estimating equations.

Proof Theorem 5: To prove Theorem 5, it suffices to show that the conditional distribution of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ converges to the limiting distribution of $n^{1/2}(\hat{\theta} - \theta_0)$, where $\hat{\theta}^T = (\hat{\gamma}_G^T, \hat{\beta}_M^T)$ and $\hat{\theta}^{*T} = (\hat{\gamma}_G^{*T}, \hat{\beta}_M^{*T})$ is the solution of the perturbed estimating equations. Suppose we perturb the two estimation equations with the same positive random variables $R_i(i = 1, ..., n)$ whose mean and variance are all one. Then, by the similar arguments in proof of Theorem 3, $\hat{\theta}^* \to \hat{\theta}$ almost surely.

From the results of Jin et al. (2006a), it follows that

$$n^{\frac{1}{2}}(\hat{\beta}_M^* - \hat{\beta}_M) = I^*(\beta_{zo}, \beta_0)^{-1} n^{-\frac{1}{2}} U_{2,M}^*(\hat{\beta}_{zo}, \hat{\beta}_M) + o_p(1), \tag{2.37}$$

and

$$U_{2,M}^*(\hat{\beta}_{zo}, \hat{\beta}_M) = U_{2,M}^*(\hat{\beta}_{zo}, \hat{\beta}_M) - U_{2,M}(\hat{\beta}_{zo}, \hat{\beta}_M) + o(n^{1/2}), \tag{2.38}$$

as $(\hat{\beta}_{zo}, \hat{\beta}_M)$ is a root of $U_{2,M}(\beta_o, \beta)$. From (2.38), it follows that

$$n^{-\frac{1}{2}} U_{2,M}^*(\hat{\beta}_{zo}, \hat{\beta}_M) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\{\tilde{Z}_i - \Phi(\hat{\beta}_{zo} + \hat{\beta}_M^T D_i)\}\phi(\hat{\beta}_{zo} + \hat{\beta}_M^T D_i)}{\Phi(\hat{\beta}_{zo} + \hat{\beta}_M^T D_i)\{1 - \Phi(\hat{\beta}_{zo} + \hat{\beta}_M^T D_i)\}} D_i(R_i - 1) + o(1). \tag{2.39}$$

Conditionally on the data, the right-hand side of (2.35) and (2.39) are the sum of independent zero-mean random vectors. Therefore, the multivariate central limit theorem implies that the conditional distribution of $n^{-1/2}\{U^*_{1,G}(\hat{\gamma}_G)^T, U^*_{2,M}(\hat{\beta}_M)^T\}^T$ given the data converges to a mean 0 multivariate normal distribution with the covariance matrix

$$\tilde{\Sigma}_{GM} = \begin{pmatrix} \tilde{\Sigma}_{1,G} & \tilde{\Sigma}_{G,M} \\ \tilde{\Sigma}^T_{G,M} & \tilde{\Sigma}_{2,M} \end{pmatrix},$$

where

$$\tilde{\Sigma}_{2,M} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} U_{2,M(i)}(\hat{\beta}_{zo}, \hat{\beta}_M) U_{2,M(i)}(\hat{\beta}_{zo}, \hat{\beta}_M)^T,$$
$$\tilde{\Sigma}_{G,M} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} U_{1,G(i)}(\hat{\gamma}_G) U_{2,M(i)}(\hat{\beta}_{zo}, \hat{\beta}_M)^T,$$

and

$$U_{2,M(i)}(\hat{\beta}_{zo}, \hat{\beta}_M) = \frac{\{\tilde{Z}_i - \Phi(\hat{\beta}_{zo} + \hat{\beta}^T_M D_i)\}\phi(\hat{\beta}_{zo} + \hat{\beta}^T_M D_i)}{\Phi(\hat{\beta}_{zo} + \hat{\beta}^T_M D_i)\{1 - \Phi(\hat{\beta}_{zo} + \hat{\beta}^T_M D_i)\}} D_i.$$

As $S^{(0)}(\gamma; t) \to s^{(0)}(\gamma; t)$, $S^{(1)}(\gamma; t) \to s^{(1)}(\gamma; t)$ and $\hat{\theta} \to \theta_0$, we have $\tilde{\Sigma}_{GM} \to \Sigma_{GM}$. Then it follows from (2.32) and (2.37) that the conditional distribution of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ converges to a mean 0 multivariate normal distribution with the covariance matrix Σ_{GM}, which is the limiting distribution of $n^{1/2}(\hat{\theta} - \theta_0)$.

This appendix gives proofs of the Corollary 4 and Theorem 4. First we impose the following regularity conditions.

Condition 5. $\sup_i \|D_i\| \le M$, *where* M *is a positive constant, and* $\theta_0 \in B_{p+q}(0, \rho)$, *a* $p+q$*-dimensional ball in* R^{p+q} *centered at zero and with radius* ρ. *In addition,* $\mu_i = \beta_0^T D_i$ *has a differentiable and bounded density function* $f_\mu(\cdot)$ *and* $\sigma(\cdot)$ *is differentiable.*

Condition 6. *For all* ν, $F_{\theta_0}(u \mid \nu)$ *has a bounded twice-differentiable density* $f_{\theta_0}(u \mid \nu)$. *In addition,* $\int_{-\infty}^{\infty} u^2\, dF_{\theta_0}(u \mid \nu) < \infty$ *and* $\int_{-\infty}^{\infty} \{\dot{f}_{\theta_0}(u \mid \nu)\}^2\, f_{\theta_0}(u \mid \nu) du < \infty$, *where* $\dot{f}_{\theta_0}(u \mid \nu)$ *is the first derivative of* $f_{\theta_0}(u \mid \nu)$ *with respect to* u.

Condition 7. *The bandwidth satisfies* $h_n = O(n^{-1/2+\kappa})$, *where* $0 < \kappa \le 1/6$.

Condition 8. *The kernel function* $K(\cdot)$ *is Lipschitz continuous of order one and satisfies* $\int K(u)du = 1$, $\int uK(u)du = 0$, $\int K^2(u)du < \infty$ *and* $\int u^2 K(u)du < \infty$.

Condition 9. *There exist some constants* e_1 *and* $e_2 > 0$ *such that* $P(C_i^Y - \gamma_0^T D_i > e_1) = 0$ *and* $\inf_\nu F_{\theta_0}(e_1 \mid \nu) > e_2$ *for all* ν.

Condition 10. *For* $0 < \lambda < \frac{1}{12}$, $\lim_{n\to\infty} n^{-3/4}\{\inf_{\|\gamma\|\le\rho, \|\gamma-\gamma_0\|\ge n^{-\lambda}} \|V_{1,B}(\gamma)\|\} = \infty$.

Condition 11. *The first order derivative matrix* Γ_n *of* $n^{-1}V_{1,B}(\gamma)$ *at* γ_0 *converges to a finite and nondegenerate matrix* Γ, *as* n *goes to infinity.*

Condition 12. *There exist positive finite constants* c_1 *and* c_2 *such that for all* i, $E(| \delta_i^{*2} |^{1+c_1}) < c_2$ *and* $\bar{M} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n D_{i(c)} D_{i(c)}^T$ *is nonsingular.*

Condition 13. *The limit of the average covariance matrix*
$\bar{\Delta} = lim_{n\to\infty} n^{-1} \sum_{i=1}^n \delta_i^{*2} D_{i(c)} D_{i(c)}^T$ *is nonsingular.*

Proof Corollary 4. Under Condition 5-13, by Theorem 1 of Pang et al. (2014) and Lemma 1 of White (1980), $\hat{\gamma}_B \to \gamma_0$ and $\hat{\beta}_P \to \beta_0$ almost surely.

Pang et al. (2014) showed that under Condition 5-11 for sufficiently large n,

$$n^{\frac{1}{2}}(\hat{\gamma}_{LBJ} - \gamma_0) = n^{-\frac{1}{2}}\Gamma^{-1}U_{1,B}(\gamma_0) + o_p(1), \tag{2.40}$$

and that

$$n^{-\frac{1}{2}}U_{1,B}(\gamma_0) \to N(0, \Sigma_{1,B}), \text{ in distribution}, \tag{2.41}$$

where

$$U_{1,B}(\gamma_0) = \sum_{i=1}^n (D_i - \bar{D})\{\gamma_0^T D_i - \hat{Y}_i(\gamma_0)\}$$

$$= \sum_{i=1}^{n} \left\{ \int_{-\infty}^{\infty} t \, dY_i^D(t,\gamma_0) + \int_{-\infty}^{\infty} \int_t^{\infty} \frac{1 - \hat{F}_{i\theta_0}(s)}{1 - \hat{F}_{i\theta_0}(t)} ds dJ_i^D(t,\gamma_0) \right\},$$

$Y_i^D(t,\gamma_0) = (D_i - \bar{D}) I\{e_i(\gamma_0) \geq t\}$ and $J_i^D(t,\gamma_0) = (D_i - \bar{D}) I\{e_i(\gamma_0) \geq t, \Delta_i^Y = 0\}$, and $\hat{F}_{i\theta_0}(t)$ is the shorthand notation of $\hat{F}_{\theta_0}(t \mid \beta_0^T D_i)$, where $\theta_0^T = (\gamma_0^T, \beta_0^T)$.

To prove the asymptotic normality of $n^{-1/2} U_{1,B}(\gamma_0)$, Pang et al. (2014) used independent and identically distributed representation of local Kaplan-Meier estimators (Gonzalez-Manteiga and Cadarso-Suarez 1994). Let $Y_i(t) = I\{e_i(\gamma_0) \geq t\}$ and $J_i(t) = I\{e_i(\gamma_0) \geq t, \Delta_i^Y = 0\}$. Define

$$\begin{aligned}
u_{B1}(\gamma_0) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} (D_i - \bar{D}) \left\{ \int_{-\infty}^{\infty} t \, dY_i(t) + \int_{-\infty}^{\infty} \int_t^{\infty} \frac{1 - F_{i\theta_0}(s)}{1 - F_{i\theta_0}(t)} ds dJ_i(t) \right\}, \\
u_{B2}(\gamma_0) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} (D_i - \bar{D}) \int_{-\infty}^{\infty} \int_t^{\infty} \frac{1 - F_{i\theta_0}(s)}{\{1 - F_{i\theta_0}(t)\}^2} \{\hat{F}_{i\theta_0}(t) - F_{i\theta_0}(t)\} ds dJ_i(t), \\
u_{B3}(\gamma_0) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} (D_i - \bar{D}) \int_{-\infty}^{\infty} \int_t^{\infty} \frac{\hat{F}_{i\theta_0}(s) - F_{i\theta_0}(s)}{1 - F_{i\theta_0}(t)} ds dJ_i(t).
\end{aligned}$$

Further define

$$A_i = \int_{-\infty}^{\infty} t \, dY_i(t) + \int_{-\infty}^{\infty} \int_t^{\infty} \frac{1 - F_{i\theta_0}(s)}{1 - F_{i\theta_0}(t)} ds dJ_i(t).$$

Based on the consistency of $\hat{F}_{i\theta_0}(t)$, we have

$$n^{-\frac{1}{2}} U_{1,B}(\gamma_0) = u_{B1}(\gamma_0) + u_{B2}(\gamma_0) + u_{B3}(\gamma_0) + o_p(1).$$

Pang et al. (2014) showed that for $i = 1, ...n$, $E(A_i \mid D_i)$ has a common value denoted by μ_A. Let $\mu_D = E(D_i)$, $\mu_i = \beta_0^T D_i$ and $G_{\gamma_0}(t \mid D_i) = P(C_i^Y - \gamma_0^T D_i > t \mid D_i)$. It follows from Theorem 2.3 of Gonzalez-Manteiga and Cadarso-Suarez (1994) that

$$\hat{F}_{i\theta_0}(t) - F_{i\theta_0}(t) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{\mu_j - \mu_i}{h_n}\right) \xi(\tilde{Y}_j, \Delta_j^Y, t, \mu_i) + O_p(n^{-\frac{1}{2}+\varepsilon}), \tag{2.42}$$

where conditional on D_i, $\xi(\tilde{Y}_j, \Delta_j^Y, t, \mu_i)$, $j = 1, ..., n$, are independent random variables with zero mean and finite variance for any t. Plugging the independent and identically distributed representation of (2.42) into $u_{B2}(\gamma_0)$ and $u_{B3}(\gamma_0)$ gives the following result,

$$n^{-\frac{1}{2}}U_{1,B}(\gamma_0) = n^{-\frac{1}{2}}\sum_{i=1}^{n} u_{1,B(i)}(\gamma_0) + o_p(1),$$

where $u_{1,B(i)}(\gamma_0) = (D_i - \mu_D)(A_i - \mu_A) + \eta(\tilde{Y}_i, \Delta_i^Y, \mu_i) + \varphi(\tilde{Y}_i, \Delta_i^Y, \mu_i)$ and

$$\eta(\tilde{Y}_i, \Delta_i^Y, \mu_i) =$$
$$E_D\left\{(D_i - \mu_D)f_\mu(\mu_i)\int_{-\infty}^{\infty}\frac{\xi(\tilde{Y}_i, \Delta_i^Y, t, \mu_i)}{1 - F_{\theta_0}(t \mid \mu_i)}\int_t^{\infty}\{1 - F_{\theta_0}(s \mid \mu_i)\}dsdG_{\gamma_0}(t \mid D_i)\middle|\beta_0^T D_i = \mu_i\right\},$$
$$\varphi(\tilde{Y}_i, \Delta_i^Y, \mu_i) = E_D\left\{(D_i - \mu_D)f_\mu(\mu_i)\int_{-\infty}^{\infty}\xi(\tilde{Y}_i, \Delta_i^Y, s, \mu_i)dsdG_{\gamma_0}(t \mid D_i)\middle|\beta_0^T D_i = \mu_i\right\},$$

The asymptotic covariance matrix of $n^{-1/2}U_{1,B}(\gamma_0)$ is denoted as $\Sigma_{1,B}$ and multivariate central limit theorem implies (2.41).

We can write $n^{1/2}(\hat{\beta}_P - \beta_0)$ as

$$n^{\frac{1}{2}}(\hat{\beta} - \beta_0) = n^{-\frac{1}{2}}\bar{M}^{-1}U_{2,P}(\beta_0) + o_p(1). \tag{2.43}$$

Under Condition 12-13, by Lemma 2 of White (1980), the asymptotic variance of $n^{-1/2}U_{2,P}(\beta_0)$ is $\bar{\Delta}$.

We assume that the limit of the covariance matrix of $n^{-1/2}u_{1,B}(\gamma_0)$ and $n^{-1/2}U_{2,P}(\beta_0)$

$$\Sigma_{B,P} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} u_{1,B(i)}(\gamma_0)U_{2,P(i)}(\beta_0)^T \tag{2.44}$$

is non-singular where $U_{2,P(i)}(\beta_0) = \sum_{i=1}^{n}(D_i - \bar{D})(Z_i - D_i^T\beta_0)$.

In the view of (2.40), (2.41), (2.43) and (2.44), multivariate central theorem implies that $n^{1/2}\{(\hat{\gamma}_B, \hat{\beta}_P) - (\gamma_0, \beta_0)\}$ converges to a mean-zero normal distribution with the covariance matrix

$$\Sigma_{BP} = \begin{pmatrix} \Gamma^{-1}\Sigma_{1,B}(\Gamma^{-1})^T & \Gamma^{-1}\Sigma_{B,P}\bar{M}^{-1} \\ \bar{M}^{-1}\Sigma_{B,P}^T\Gamma^{-1} & \bar{M}^{-1}\bar{\Delta}\bar{M}^{-1} \end{pmatrix}.$$

□

Proof Theorem 6: To prove Theorem 6, it suffices to show that the conditional distribution of $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ converges to the limiting distribution of $n^{1/2}(\hat{\theta} - \theta_0)$, where $\hat{\theta}^T = (\hat{\gamma}_B^T, \hat{\beta}_P^T)$ and $\hat{\theta}^{*T} = (\hat{\gamma}_B^{*T}, \hat{\beta}_P^{*T})$ is the solution of the perturbed estimating equations.

The perturbed Buckley-James estimator is the solution to

$$U_{1,B}^*(\gamma) = \sum_{i=1}^n R_i(D_i - \bar{D}_n)\{\hat{Y}_i^*(\gamma) - \bar{Y}^*(\gamma) - \gamma^T(D_i - \bar{D})\}.$$

Following the proofs of Jin et al. (2006b) and Pang et al. (2014), we incorporate the random weights R_i into the derivation of the asymptotic linearity in Pang et al. (2014). As a result,

$$n^{\frac{1}{2}}(\hat{\gamma}_B^* - \hat{\gamma}_B) = n^{-\frac{1}{2}}\Gamma^{-1}U_{1,B}^*(\hat{\gamma}_B) + o_p(1). \tag{2.45}$$

Since $\hat{\gamma}_B$ is the solution to $U_{1,B}(\gamma)$,

$$U_{1,B}^*(\hat{\gamma}_B) = U_{1,B}^*(\hat{\gamma}_B) - U_{1,B}(\hat{\gamma}_B) + o(n^{\frac{1}{2}}). \tag{2.46}$$

Clearly,

$$\begin{aligned} U_{1,B}^*(\hat{\gamma}_B) - U_{1,B}(\hat{\gamma}_B) = &\sum_{i=1}^n (R_i - 1)(D_i - \bar{D})\{\hat{Y}_i^*(\hat{\gamma}_B) - \bar{Y}^*(\hat{\gamma}_B) - \hat{\gamma}_B^T(D_i - \bar{D})\} \\ &+ \sum_{i=1}^n (D_i - \bar{D})\{\hat{Y}_i^*(\hat{\gamma}_B) - \hat{\gamma}_B^T D_i\} - U_{1,B}(\hat{\gamma}_B) \end{aligned} \tag{2.47}$$

Since $E(R_i - 1|\tilde{Y}_i, \Delta_i^Y, D_i) = 0$ and $\hat{Y}_i^*(\hat{\gamma}_B) - \hat{\gamma}_B^T D_i$ can be approximated by $E(Y_i|\tilde{Y}_i, \Delta_i^Y, D_i) -$

$\gamma_0^T D_i$,

$$
\begin{aligned}
&\sum_{i=1}^{n}(R_i - 1)(D_i - \bar{D})\{\hat{Y}_i^*(\hat{\gamma}_B) - \bar{Y}^*(\hat{\gamma}_B) - \hat{\gamma}_B^T(D_i - \bar{D})\} \\
=&\sum_{i=1}^{n}(R_i - 1)(D_i - \bar{D})\{E(Y_i|\tilde{Y}_i, \Delta_i^Y, D_i) - \gamma_0^T D_i\} + o(n^{\frac{1}{2}}) \\
=&\sum_{i=1}^{n}(R_i - 1)(D_i - \bar{D})\left\{\int_{-\infty}^{\infty} t\, dY_i(t) + \int_{-\infty}^{\infty}\int_{t}^{\infty} \frac{1 - F_{i\theta_0}(s)}{1 - F_{i\theta_0}(t)} ds dJ_i(t)\right\} + o(n^{\frac{1}{2}}) \\
=&\sum_{i=1}^{n}(R_i - 1)(D_i - \mu_D)(A_i - \mu_A) + o_p(n^{\frac{1}{2}}).
\end{aligned} \tag{2.48}
$$

Note that

$$
\begin{aligned}
&\sum_{i=1}^{n}(D_i - \bar{D})\{\hat{Y}_i^*(\hat{\gamma}_B) - \hat{\gamma}_B^T D_i\} - U_{1,B}(\hat{\gamma}_B) \\
=&\sum_{i=1}^{n}(D_i - \bar{D})\int_{-\infty}^{\infty}\int_{t}^{\infty} \frac{1 - \hat{F}_{i\hat{\theta}}(s)}{\{1 - \hat{F}_{i\hat{\theta}}(t)\}^2}\{\hat{F}_{i\hat{\theta}}^*(t) - \hat{F}_{i\hat{\theta}}(t)\} ds dJ_i(t) \qquad (2.49) \\
&+ \sum_{i=1}^{n}(D_i - \bar{D})\int_{-\infty}^{\infty}\int_{t}^{\infty} \frac{\hat{F}_{i\hat{\theta}}^*(s) - \hat{F}_{i\hat{\theta}}(s)}{1 - \hat{F}_{i\hat{\theta}}(t)} ds dJ_i(t) \qquad (2.50)
\end{aligned}
$$

As in Jin et al. (2006b), by approximating $\hat{F}_{i\hat{\theta}}^*(t) - \hat{F}_{i\hat{\theta}}(t)$ with a weighted sum of $R_j - 1$, $j = 1, ..., n$, (2.49) can be written as

$$
\sum_{i=1}^{n}(R_i - 1)\eta(\tilde{Y}_i, \Delta_i^Y, \mu_i) + \sum_{i=1}^{n}(R_i - 1)\varphi(\tilde{Y}_i, \Delta_i^Y, \mu_i) + o_p(n^{\frac{1}{2}}) \tag{2.51}
$$

From (2.46), (2.48) and (2.51), it follows that

$$
n^{-\frac{1}{2}} U_{1,B}^*(\hat{\gamma}_B) = n^{-\frac{1}{2}} \sum_{i=1}^{n}(R_i - 1) u_{1,B(i)}(\gamma_0) + o_p(1). \tag{2.52}
$$

The perturbed least squares estimator denoted by $\hat{\beta}^*$ is the solution to

$$
U_{2,P}^*(\beta) = \sum_{i=1}^{n} R_i(D_i - \bar{D})(Z_i - D_i^T\beta).
$$

By the same arguments used in previous theorems and approximating $\hat{\beta}_P$ by β_0, we have

$$n^{-\frac{1}{2}} U^*_{2,P}(\beta) = n^{-\frac{1}{2}} \sum_{i=1}^{n} (R_i - 1)(D_i - \bar{D})(Z_i - \beta_0^T D_i) + o_p(1). \tag{2.53}$$

Conditionally on the data, the right-hand side of (2.52) and (2.53) are the sum of independent zero-mean random vectors. Therefore, the multivariate central limit theorem implies that the conditional distribution of $n^{-1/2}\{U^*_{1,B}(\hat{\gamma}_B)^T, U^*_{2,P}(\hat{\beta}_P)^T\}^T$ given the data converges to a mean 0 multivariate normal distribution with the covariance matrix Σ_{BP}, which is the limiting distribution of $n^{1/2}(\hat{\theta} - \theta_0)$.

Now we will show that the converting the estimate of the latent exposure in Case 2-1 analysis to that of binary exposure does not give a correct result in general. Remind that our true response model is given by

$$Y_i = \alpha_{yo} + \alpha_{yz} Z_i + \alpha_{yw}^T W_i + \varepsilon_i^*, \tag{2.54}$$

where $E(\varepsilon_i^* \mid V_i, W_i) = 0$. Consider the linear model with binary exposure,

$$Y_i = \tilde{\alpha}_{yo} + \tilde{\alpha}_{yz} \tilde{Z}_i + \tilde{\alpha}_{yw}^T W_i + \tilde{\varepsilon}_i^*, \tag{2.55}$$

where $\tilde{Z}_i = I(Z_i > 0)$. Note that the parameters in (2.55) may be different from those in (2.54). Suppose we want to obtain the estimate of $\tilde{\alpha}_{yz}$ from the estimates of the parameters in model (2.54). As in model (2.54), we impose the conditional independence assumption of the instrumental variables on $\tilde{\varepsilon}_i^*$, $E(\tilde{\varepsilon}_i^* \mid V_i, W_i) = 0$.

Note that

$$\begin{aligned} \tilde{\varepsilon}_i^* &= Y_i - \tilde{\alpha}_{yo} - \tilde{\alpha}_{yz} \tilde{Z}_i - \tilde{\alpha}_{yw}^T W_i \\ &= \varepsilon_i^* + (\alpha_{yo} - \tilde{\alpha}_{yo}) + \alpha_{yz} Z_i - \tilde{\alpha}_{yz} \tilde{Z}_i + (\alpha_{yw} - \tilde{\alpha}_{yw})^T W_i. \end{aligned}$$

The equation $E(\tilde{\varepsilon}_i^* \mid D_i) = 0$ can be written as

$$E(\varepsilon_i^* \mid D_i) + (\alpha_{yo} - \tilde{\alpha}_{yo}) + \alpha_{yz} E(Z_i \mid D_i) - \tilde{\alpha}_{yz} E(\tilde{Z}_i \mid D_i) + (\alpha_{yw} - \tilde{\alpha}_{yw})^T W_i = 0. \tag{2.56}$$

From (2.56), it follows that

$$\tilde{\alpha}_{yz} E(\tilde{Z}_i \mid D_i) = E(\varepsilon_i^* \mid D_i) + (\alpha_{yo} - \tilde{\alpha}_{yo}) + \alpha_{yz} E(Z_i \mid D_i) + (\alpha_{yw} - \tilde{\alpha}_{yw})^T W_i.$$

Since $E(\varepsilon_i^* \mid D_i) = 0$, we have

$$\tilde{\alpha}_{yz} = \frac{\alpha_{yo} - \tilde{\alpha}_{yo}}{n^{-1} \sum_{i=1}^n E(\tilde{Z}_i \mid D_i)} + \alpha_{yz} \frac{\sum_{i=1}^n E(Z_i \mid D_i)}{\sum_{i=1}^n E(\tilde{Z}_i \mid D_i)} + \frac{(\alpha_{yw} - \tilde{\alpha}_{yw})^T \sum_{i=1}^n W_i}{\sum_{i=1}^n E(\tilde{Z}_i \mid D_i)}.$$

If $\alpha_{yo} = \tilde{\alpha}_{yo}$ and $\alpha_{yw} = \tilde{\alpha}_{yw}$, then $\tilde{\alpha}_{yz}$ reduces to

$$\tilde{\alpha}_{yz} = \alpha_{yz} \frac{\sum_{i=1}^n E(Z_i \mid D_i)}{\sum_{i=1}^n E(\tilde{Z}_i \mid D_i)},$$

hence we can estimate $\tilde{\alpha}_{yz}$ from the estimate of α_{yz}, $E(Z_i \mid D_i)$ and $E(\tilde{Z}_i \mid D_i)$. However, in general, $\alpha_{yo} = \tilde{\alpha}_{yo}$ and $\alpha_{yw} = \tilde{\alpha}_{yw}$ does not hold.

CHAPTER 3: A NEW INSTRUMENTAL VARIABLE ESTIMATOR USING A NEGATIVE CONTROL OUTCOME

3.1 Introduction

Confounders are the variables that affect both exposure and outcome. Regression adjustment and propensity score method are usually used to rule out the bias by confounding. However, it is hard to guarantee that all of the confounders are measured. If such an unmeasured confounder exists, then the aforementioned methods that adjust for the confounders do not give unbiased estimates of the exposure effect.

Instrumental variable (IV) methods are designed to eliminate the bias by unmeasured confounding, thus they give unbiased estimates of the exposure effect. The variable V is called a valid IV if it satisfies the following three assumptions (Brookhart et al. 2010): (i) V has an effect on the exposure Z, (ii) V has an effect on the outcome Y only through Z, (iii) V is unrelated to an unmeasured confounder U. In terms of assumption (i), the strength of the association between Z and V determines the validity of the IV estimator: a strong association yields an efficient and unbiased IV estimator. The assumption (ii) is called the exclusion restriction (Angrist et al. 1996) and is needed to identify the estimand for the exposure effect via an IV (Angrist et al. 1996). The assumption (iii) is satisfied if U is well balanced between the IV groups when the IV is binary. We will call assumption (iii) IV independence assumption.

Negative controls are useful tools to detect unmeasured confounding (Flanders et al. 2011, Lipsitch et al. 2010, Lumley and Sheppard 2000, Jackson et al. 2006, Smith 2008; 2012). An outcome N is called a valid negative control outcome if N satisfies the following two conditions (Lipsitch et al. 2010, Tchetgen Tchetgen 2014): (i) N shares a set of common confounders of Z and Y, (ii) N is not caused by Z conditional on the common confounders. Those conditions are sufficient to identify the existence of unmeasured confounding, but insufficient to obtain the unbiased estimate of the exposure effect. Tchetgen Tchetgen (2014) proposed control outcome calibration methods to estimate the exposure effect by imposing an assumption that N is independent of Z conditional on measured confounders W and counterfactual outcomes. This assumption holds when the counterfactual outcomes are ideal proxies of U (the variation of U is completely explained by that of the outcomes), however, this may not happen in practice because the measurement of the outcomes usually are exposed to some errors.

In observation studies, it may not be easy to find a valid IV that satisfies all the three IV assumptions. The weak IV problems have been focused well and theoretical results of IV models with weak IVs have been studied by some researchers (Staiger and Stock 1997, Stock and Yogo 2005). However, to our knowledge, there is a lack of literatures to deal with the violation of the IV independence assumption.

In Section 2, we propose a new IV estimator using a negative control outcome. The new IV estimator has been shown to be theoretically valid even if the IV independence assumption is violated. Structural equation models are used to define the IV and the negative control outcome. For the case of multiple IVs, the new IV method depends on weights and the optimal weight is identified. In section 3, a Wald test is proposed to test the IV independence assumption and a new IV estimator combining the new IV estimator and an usual IV estimator is developed, which may be superior to both individual IV estimators. The new IV estimator is shown to perform well in simulations reported in Section 4.

3.2 A new instrumental variable estimator with a negative control outcome

We start with a simple case where one instrumental variable is available. For $i = 1, ..., n$, suppose Y_i is the main outcome, N_i is the negative control outcome, Z_i is the exposure, V_i is the instrumental variable, $W_i = (w_{i1}, ..., w_{iq})^T$ is the $q \times 1$ vector of (measured) confounders and U_i is the unmeasured confounder.

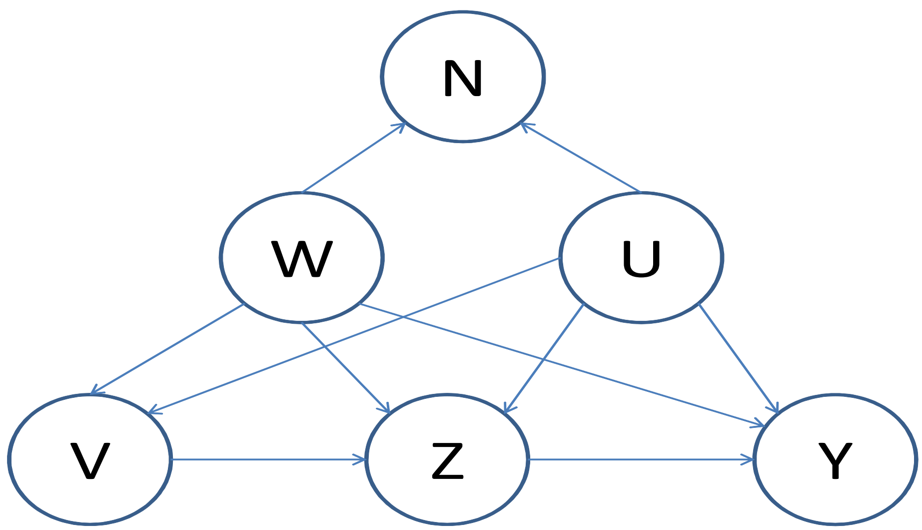

Figure 3.1: Causal diagram showing an ideal negative control outcome N for use in estimating an exposure effect on an outcome with an instrument V. N should have the same incoming arrow as an outcome Y, except that N is not caused by Z. V can be caused by measured and unmeasured confounders W and U.

For the main and negative control outcomes, we consider the following linear models:

$$Y_i = \alpha_{yo} + \alpha_{yz} Z_i + \alpha_{yv} V_i + \alpha_{yw}^T W_i + \alpha_{yu} U_i + \varepsilon_i, \tag{3.1}$$

$$N_i = \alpha_{no} + \alpha_{nz} Z_i + \alpha_{nv} V_i + \alpha_{nw}^T W_i + \alpha_{nu} U_i + \xi_i, \tag{3.2}$$

$\alpha_{yw} = (\alpha_{yw,1}, ..., \alpha_{yw,q})^T$ is a $q \times 1$ parameter vector and $E(\varepsilon_i | X_i, U_i) = E(\xi_i | X_i, U_i) = 0$ for all i by construction and $X_i^T = (1, Z_i, V_i, W_i^T)$. Assume that $(X_i, U_i, \varepsilon_i, \xi_i)$ is a sequence of independent and not (necessarily) identically distributed random vectors. The conditional covariance matrix of (ε_i, ξ_i) given X_i is allowed to depend on X_i. In (3.1), we set $\alpha_{yv} = 0$ as V_i and Y_i are conditionally independent given (Z_i, W_i^T, U_i), that is, the exclusion restriction assumption (Angrist et al., 1996). In (3.2), we also set $\alpha_{nv} = 0$ by the exclusion restriction assumption and $\alpha_{nz} = 0$ following the definition of the negative control outcome, that is, there is no effect of Z_i on N_i conditional on (W_i^T, U_i).

We consider the situation where U_i is unmeasured. In vector notation, the models of (3.1) and (3.2) can be written as

$$Y = X\alpha_y + \alpha_{yu} U + \varepsilon, \tag{3.3}$$

$$N = X\alpha_n + \alpha_{nu} U + \xi, \tag{3.4}$$

where $\alpha_y^T = (\alpha_{yo}, \alpha_{yz}, \alpha_{yv}, \alpha_{yw}^T)$, $\alpha_n^T = (\alpha_{no}, \alpha_{nz}, \alpha_{nv}, \alpha_{nw}^T)$, $Y = (Y_1, ..., Y_n)^T$, $N = (N_1, ..., N_n)^T$, $U = (U_1, ..., U_n)^T$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)^T$, $\xi = (\xi_1, ..., \xi_n)^T$ and X is the matrix with a typical row X_i.

The least squares estimators for α_y and α_n with observed covariates are given by

$$\hat{\alpha}_y = (X^T X)^{-1} X^T Y = \alpha_y + (X^T X)^{-1} X^T \varepsilon + \alpha_{yu} (X^T X)^{-1} X^T U, \tag{3.5}$$

$$\hat{\alpha}_n = (X^T X)^{-1} X^T N = \alpha_n + (X^T X)^{-1} X^T \xi + \alpha_{nu} (X^T X)^{-1} X^T U. \tag{3.6}$$

In general, $n^{-1} X^T U$ dose not converge to zero because Z_i and U_i are correlated. Therefore, $\hat{\alpha}_y$ is not a consistent estimator for α_y.

Let $S_i^T = (X_i^T, U_i)$ and $r = 4+q$ denote the length of S_i. Further let S_{ij} denote the jth element of the vector S_i. To identify the limiting values of $\hat{\alpha}_y$ and $\hat{\alpha}_n$, we make the following assumption.

Assumption 3. *(a) There exist positive finite constants δ and Δ such that, for all i, $E(|\varepsilon_i^2|^{1+\delta}) < \Delta$, $E(|\xi_i^2|^{1+\delta}) < \Delta$ and $E(|S_{ij}S_{ik}|^{1+\delta}) < \Delta(j,k = 1,...,r)$; (b) $\bar{M}_n = n^{-1}E(X^T X)$ is nonsingular for (all) n sufficiently large.*

Under Assumption 3, it follows by strong law of large numbers that almost surely

$$(X^T X)^{-1}X^T\varepsilon \to 0,\ (X^T X)^{-1}X^T\xi \to 0,\ (X^T X)^{-1}X^T U \to \bar{M}_n^{-1}n^{-1}E(X^T U). \tag{3.7}$$

Denote $\bar{M}_n^{-1}n^{-1}E(X^T U)$ by $\nu = (\nu_o, \nu_z, \nu_v, \nu_w^T)^T$. From (3.7), it follows that

$$\hat{\alpha}_y = \alpha_y + \alpha_{yu}\nu + o(1) = \alpha_y^* + o(1), \tag{3.8}$$

$$\hat{\alpha}_n = \alpha_n + \alpha_{nu}\nu + o(1) = \alpha_n^* + o(1), \tag{3.9}$$

where $\alpha_y^* = \alpha_y + \alpha_{yu}\nu$ and $\alpha_n^* = \alpha_n + \alpha_{nu}\nu$. Note that $\hat{\alpha}_y$ and $\hat{\alpha}_n$ are biased estimators for α_y and α_n, but unbiased estimators for α_y^* and α_n^*.

The proposed instrumental variable estimator for α_{yz} is given by

$$\hat{\alpha}_{yz}^{NC} = \hat{\alpha}_{yz} - \frac{\hat{\alpha}_{yv}}{\hat{\alpha}_{nv}}\hat{\alpha}_{nz}, \tag{3.10}$$

where $\hat{\alpha}_{nv}$ is non-zero. Using (3.8), (3.9) and continuous mapping theorem, we can show that the proposed estimator converses almost surely to α_{yz},

$$\begin{aligned}
\hat{\alpha}_{yz}^{NC} &= (\alpha_{yz} + \alpha_{yu}\nu_z) - \frac{\alpha_{yv} + \alpha_{yu}\nu_v}{\alpha_{nv} + \alpha_{nu}\nu_v}(\alpha_{nz} + \alpha_{nu}\nu_z) + o(1),\\
&= (\alpha_{yz} + \alpha_{yu}\nu_z) - \frac{\alpha_{yu}\nu_v}{\alpha_{nu}\nu_v}\alpha_{nu}\nu_z + o(1),\\
&= \alpha_{yz} + o(1).
\end{aligned}$$

The critical condition for which the proposed instrumental variable estimator is valid is that both α_{nu} and ν_v are non-zeros because the estimator is a function of $1/\hat{\alpha}_{nv}$ and the estimator $\hat{\alpha}_{nv}$ converges to $\alpha_{nu}\nu_v$. The condition of $\alpha_{nu} \neq 0$ indicates that N_i shares W_i and U_i with Y_i. This is called "U-comparable" condition for the use of the negative control outcome to detect unmeasured confounding (Lipsitch et al. 2010). The condition of $\nu_v \neq 0$ implies the correlation between V_i and U_i and this condition holds if both V_i and U_i affect Z_i or there is a direct relationship between V_i and U_i. Thus the proposed new instrumental variable method is reliable even if an usual instrumental variable assumption that V_i is independent of U_i is violated.

Next we consider multiple instrumental variables. Define $V_i = (V_{i1}, ..., V_{ip})^T$ as the $p \times 1$ vector of instruments. Further Define $X_i^T = (1, Z_i, V_i^T, W_i^T)$, $\alpha_{yv} = (\alpha_{yv,1}, ..., \alpha_{yv,p})^T$, $\alpha_y^T = (\alpha_{yo}, \alpha_{yz}, \alpha_{yv}^T, \alpha_{yw}^T)$, $\alpha_n^T = (\alpha_{no}, \alpha_{nz}, \alpha_{nv}^T, \alpha_{nw}^T)$ and ν^T=(ν_o,ν_z,ν_v^T, ν_w^T). In the case of multiple instrumental variables, the proposed instrumental variable estimator for α_{yz} has the following form,

$$\begin{aligned}\hat{\alpha}_{yz}^{NC} &= \hat{\alpha}_{yz} - \left(\omega_n^T \frac{\hat{\alpha}_{yv}}{\hat{\alpha}_{nv}}\right)\hat{\alpha}_{nz}, \\ &= \hat{\alpha}_{yz} - \left(\sum_{j=1}^{p} \omega_{nj} \frac{\hat{\alpha}_{yv,j}}{\hat{\alpha}_{nv,j}}\right)\hat{\alpha}_{nz}, \end{aligned} \tag{3.11}$$

where the division sign in the first line is understood to perform element by element and $\omega_n = (\omega_{n1}, ..., \omega_{np})^T \to \omega_0 = (\omega_1, ..., \omega_p)^T$ almost surely with the condition of $\omega_0^T 1_p = \sum_{j=1}^p \omega_j = 1$, where 1_p is the $p \times 1$ vector of ones. In the sequel, the division sign used for vectors is understood to indicate division of two vectors element by element. The following theorem states the strong consistency of $\hat{\alpha}_{yz}^{NC}$.

Theorem 7. *Given Assumption 3 and ω_0 such that $\sum_{j=1}^p \omega_j = 1$, $\hat{\alpha}_{yz}^{NC}$ converges almost surely to α_{yz}.*

Proof. By continuous mapping theorem, it follows that

$$\hat{\alpha}_{yz}^{NC} = (\alpha_{yz} + \alpha_{yu}\nu_z) - \left(\omega_0^T \frac{\alpha_{yu}\nu_v}{\alpha_{nu}\nu_v}\right)\alpha_{nu}\nu_z + o(1),$$

$$
\begin{aligned}
&= (\alpha_{yz} + \alpha_{yu}\nu_z) - (\omega_0^T 1_p)\alpha_{yu}\nu_z + o(1), \\
&= \alpha_{yz} + o(1).
\end{aligned}
$$

$\square$

Note that the parameter α_{yz} can be written as

$$
\begin{aligned}
\alpha_{yz} &= \alpha_{yz}^* - \omega_0^T \gamma_{nz}^*, \\
&= (\alpha_{yz} + \alpha_{yu}\nu_z) - \omega_0^T 1_p(\alpha_{yu}\nu_z), \qquad (3.12)
\end{aligned}
$$

where $\gamma_{nz}^* = (\alpha_{yv}^*/\alpha_{nv}^*)\alpha_{nz}^*$. Any ω_0 satisfying $\omega_0^T 1_p = 1$ makes (3.12) hold. Define the estimator with a known weight ω_0,

$$
\hat{\alpha}_{yz}^\dagger = \hat{\alpha}_{yz} - \omega_0^T \hat{\gamma}_{nz}, \qquad (3.13)
$$

where $\hat{\gamma}_{nz} = (\hat{\alpha}_{yv}/\hat{\alpha}_{nv})\hat{\alpha}_{nz}$. Note that $n^{1/2}(\hat{\alpha}_{yz}^{NC} - \alpha_{yz})$ can be written as

$$
n^{1/2}(\hat{\alpha}_{yz}^{NC} - \alpha_{yz}) = n^{1/2}(\hat{\alpha}_{yz} - \hat{\alpha}_{yz}^\dagger) + n^{1/2}(\hat{\alpha}_{yz}^\dagger - \alpha_{yz}). \qquad (3.14)
$$

and that

$$
\begin{aligned}
n^{1/2}(\hat{\alpha}_{yz} - \hat{\alpha}_{yz}^\dagger) &= -n^{1/2}(\omega_n - \omega_0)^T \hat{\gamma}_{nz} \\
&= -n^{1/2}(\omega_n - \omega_0)^T(\hat{\gamma}_{nz} - \gamma_{nz}^*) - n^{1/2}(\omega_n - \omega_0)^T \gamma_{nz}^*. \qquad (3.15)
\end{aligned}
$$

If we assume $n^{\frac{1}{2}}(\hat{\gamma}_{nz} - \gamma_{nz}^*) = O_p(1)$, then $n^{1/2}(\omega_n - \omega_0)^T(\hat{\gamma}_{nz} - \gamma_{nz}^*) = o_p(1)$. Since $(\omega_n - \omega_0)^T \gamma_{nz}^* = 0$, $n^{1/2}(\omega_n - \omega_0)^T \gamma_{nz}^* = 0$. Thus, the distribution of $n^{1/2}(\hat{\alpha}_{yz}^{NC} - \alpha_{yz})$ converges almost surely to the limiting distribution of $n^{1/2}(\hat{\alpha}_{yz}^\dagger - \alpha_{yz})$.

We can write $n^{1/2}(\hat{\alpha}^{\dagger}_{yz} - \alpha_{yz})$ as

$$n^{1/2}(\hat{\alpha}^{\dagger}_{yz} - \alpha_{yz}) = n^{1/2}(\hat{\alpha}_{yz} - \alpha^{*}_{yz}) - \omega_0^T n^{1/2}(\hat{\gamma}_{nz} - \gamma^{*}_{nz}).$$

The asymptotic covariance matrix of $n^{1/2}(\hat{\alpha}^{\dagger}_{yz} - \alpha_{yz})$ is given by

$$\Sigma(\hat{\alpha}^{\dagger}_{yz}) = \Sigma(\hat{\alpha}_{yz}) + \omega_0^T \Sigma(\hat{\gamma}_{nz})\omega_0 - 2\omega_0^T \Sigma(\hat{\alpha}_{yz}, \hat{\gamma}_{nz}),$$

where $\Sigma(\hat{\alpha}_{yz})$, $\Sigma(\hat{\gamma}_{nz})$ and $\Sigma(\hat{\alpha}_{yz}, \hat{\gamma}_{nz})$ are the asymptotic variance-covariance matrices of $n^{1/2}(\hat{\alpha}_{yz} - \alpha^{*}_{yz})$ and $n^{1/2}(\hat{\gamma}_{nz} - \gamma^{*}_{nz})$, and the asymptotic covariance matrix between $n^{1/2}(\hat{\alpha}_{yz} - \alpha^{*}_{yz})$ and $n^{1/2}(\hat{\gamma}_{nz} - \gamma^{*}_{nz})$. Hereafter, let $\Sigma(T_1)$ and $\Sigma(T_1, T_2)$ denote the asymptotic variance-covariance matrix of T_1 and the asymptotic covariance matrix of T_1 and T_2. It can be shown that ω_0 yielding the minimum of $\Sigma(\hat{\alpha}^{\dagger}_{yz})$ with the condition of $\omega_0^T 1_p = 1$ is given by

$$\begin{aligned}
\omega_0 &= \beta_0 + \frac{1 - \beta_0^T 1_p}{1_p^T \Sigma(\hat{\gamma}_{nz})^{-1} 1_p} \Sigma(\hat{\gamma}_{nz})^{-1} 1_p, \\
&= \beta_0 + \frac{1 - \sum_{j=1}^{p} \beta_j}{\sum_{k,l=1}^{p} \Sigma_{kl}(\hat{\gamma}_{nz})^{-1}} \Sigma(\hat{\gamma}_{nz})^{-1} 1_p, \qquad (3.16)
\end{aligned}$$

where $\beta_0 = (\beta_1, ..., \beta_p)^T = \Sigma(\hat{\gamma}_{nz})^{-1}\Sigma(\hat{\alpha}_{yz}, \hat{\gamma}_{nz})$ is the value of ω_0 yielding the minimum of $\Sigma(\hat{\alpha}^{\dagger}_{yz})$ without the condition of $\omega_0^T 1_p = 1$ and $\Sigma_{kl}(\hat{\gamma}_{nz})^{-1}$ is the (k, l)th element of $\Sigma(\hat{\gamma}_{nz})^{-1}$. We choose ω_n as the consistent estimator of ω_0 in (3.16),

$$\begin{aligned}
\omega_n &= \hat{\beta} + \frac{1 - \hat{\beta}^T 1_p}{1_p^T \hat{\Sigma}(\hat{\gamma}_{nz})^{-1} 1_p} \hat{\Sigma}(\hat{\gamma}_{nz})^{-1} 1_p, \\
&= \hat{\beta} + \frac{1 - \sum_{j=1}^{p} \hat{\beta}_j}{\sum_{k,l=1}^{p} \hat{\Sigma}_{kl}(\hat{\gamma}_{nz})^{-1}} \hat{\Sigma}(\hat{\gamma}_{nz})^{-1} 1_p, \qquad (3.17)
\end{aligned}$$

where $\hat{\beta} = \hat{\Sigma}(\hat{\gamma}_{nz})^{-1}\hat{\Sigma}(\hat{\alpha}_{yz}, \hat{\gamma}_{nz})$, and $\hat{\Sigma}(\hat{\gamma}_{nz})^{-1}$ and $\hat{\Sigma}(\hat{\alpha}_{yz}, \hat{\gamma}_{nz})$ are consistent estimators of $\Sigma(\hat{\gamma}_{nz})^{-1}$ and $\Sigma(\hat{\alpha}_{yz}, \hat{\gamma}_{nz})$.

Let $\hat{\alpha} = (\hat{\alpha}_y^T, \hat{\alpha}_n^T)^T$ and $\alpha^* = (\alpha_y^{*T}, \alpha_n^{*T})^T$. It can be shown that

$$\hat{\alpha}_y - \alpha_y^* = (X^T X)^{-1} X^T \varepsilon_u, \ \hat{\alpha}_n - \alpha_n^* = (X^T X)^{-1} X^T \xi_u, \tag{3.18}$$

where

$$\varepsilon_u = (\varepsilon_{u1}, \ldots, \varepsilon_{un})^T = Y - X\alpha_y^* = \varepsilon + \alpha_{yu}(U - X\nu), \tag{3.19}$$

$$\xi_u = (\xi_{u1}, \ldots, \xi_{un})^T = N - X\alpha_n^* = \xi + \alpha_{nu}(U - X\nu), \tag{3.20}$$

Note that $E(X^T \varepsilon_u) = 0$ because of the fact that $E(X^T U) = E(X^T X)\nu$ while $E(\varepsilon_u)$ may not be zero. Thus we use the methods of White (1980) for heteroskedasticity-consistent covariance matrix estimation. Define the following average covariance matrices

$$\bar{V}_{n,\varepsilon\varepsilon} = n^{-1} \sum_{i=1}^n E(\varepsilon_{ui}^2 X_i X_i^T), \ \bar{V}_{n,\xi\xi} = n^{-1} \sum_{i=1}^n E(\xi_{ui}^2 X_i X_i^T), \ \bar{V}_{n,\varepsilon\xi} = n^{-1} \sum_{i=1}^n E(\varepsilon_{ui}\xi_{ui} X_i X_i^T),$$

and the covariance matrix

$$\Sigma_\alpha = \begin{pmatrix} \bar{M}_n^{-1} \bar{V}_{n,\varepsilon\varepsilon} \bar{M}_n^{-1} & \bar{M}_n^{-1} \bar{V}_{n,\varepsilon\xi} \bar{M}_n^{-1} \\ \bar{M}_n^{-1} \bar{V}_{n,\varepsilon\xi} \bar{M}_n^{-1} & \bar{M}_n^{-1} \bar{V}_{n,\xi\xi} \bar{M}_n^{-1} \end{pmatrix}.$$

With the next assumption, an asymptotic normality can be obtained.

Assumption 4. *Let X_{ij} be the jth element of the vector X_i. (a) There exist positive finite constants δ and Δ such that for all i $E(|\varepsilon_{ui}^2 X_{ij} X_{ik}|^{1+\delta}) < \Delta$ and $E(|\xi_{ui}^2 X_{ij} X_{ik}|^{1+\delta}) < \Delta$ $(j, k = 1, \ldots, 2 + p + q)$; (b) Σ_α is nonsingular for n sufficiently large.*

With Assumptions 3 and 4, by Lemma 2 of White (1980), $n^{1/2}(\hat{\alpha} - \alpha^*) \to N(0, \Sigma_\alpha)$ in distribution. To obtain the asymptotic distribution of $n^{1/2}(\hat{\alpha}_{yz}^\dagger - \alpha_{yz})$, we let

$$\begin{pmatrix} \hat{\alpha}_{yz} \\ \hat{\gamma}_{nz} \end{pmatrix} = g(\hat{\alpha}) = \begin{pmatrix} g_1(\hat{\alpha}) \\ g_2(\hat{\alpha}) \end{pmatrix}. \tag{3.21}$$

The first derivative of g at α^* is given by

$$\dot{g}(\alpha^*)^T = \begin{pmatrix} 0, & 1, & 0_p^T, & 0_q^T, & 0, & 0, & 0_p^T, & 0_q^T \\ 0_p, & 0_p, & \mathrm{diag}\left(\frac{\alpha_{nz}^*}{\alpha_{nv}^*}\right), & 0_{p\times q}, & 0_p, & \frac{\alpha_{yv}^*}{\alpha_{nv}^*}, & -\mathrm{diag}\left(\frac{\alpha_{yv}^*\alpha_{nz}^*}{\alpha_{nv}^{*2}}\right), & 0_{p\times q} \end{pmatrix}, \tag{3.22}$$

where $\mathrm{diag}(A)$ is a diagonal matrix with the ith diagonal element being the ith element of vector A. By Delta method, $n^{1/2}\{(\hat{\alpha}_{yz}, \hat{\gamma}_{nz}^T)^T - (\alpha_{yz}^*, \gamma_{nz}^{*T})^T\}$ converges in distribution to $N(0, \Sigma_g)$, where $\Sigma_g = \dot{g}(\alpha^*)^T\Sigma_\alpha\dot{g}(\alpha^*)$. From this result, we can obtain $\Sigma(\alpha_{yz}^*)$, $\Sigma(\gamma_{nz}^*)$ and $\Sigma(\alpha_{yz}^*, \gamma_{nz}^*)$. Further let $\hat{\alpha}_{yz}^\dagger = h(\hat{\alpha}_{yz}, \hat{\gamma}_{nz})$. The first derivative of h at $(\alpha_{yz}^*, \gamma_{nz}^{*T})^T$ is given by $\dot{h}(\alpha_{yz}^*, \gamma_{nz}^*)^T = (1, -\omega_0^T)$. By Delta method, $n^{1/2}(\hat{\alpha}_{yz}^\dagger - \alpha_{yz})$ converges in distribution to a mean-zero multivariate normal distribution with the covariance matrix of $(1, -\omega_0^T)\Sigma_g(1, -\omega_0^T)^T$. Using this asymptotic result, we obtain asymptotic normality of the proposed estimator.

Theorem 8. *With Assumptions 3 and 4 and ω_n which converges in almost surely to ω_0, $n^{1/2}(\alpha_{yz}^{NC} - \alpha_{yz}) \to N\{0, \Sigma(\alpha_{yz}^{NC})\}$ in distribution,*
where $\Sigma(\alpha_{yz}^{NC}) = (1, -\omega_0^T)\Sigma_g(1, -\omega_0^T)^T$.

To estimate Σ_α, we use the estimators of White (1980). Define the residuals to be $\hat{\varepsilon}_u = Y - X\hat{\alpha}_y$ and $\hat{\xi}_u = N - X\hat{\alpha}_n$, which are consistent for ε_u and ξ_u. The estimators of $V_{n,\varepsilon\varepsilon}$, $V_{n,\xi\xi}$ and $V_{n,\varepsilon\xi}$ are given by

$$\hat{V}_{n,\varepsilon\varepsilon} = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{ui}^2 X_i X_i^T,\ \hat{V}_{n,\xi\xi} = n^{-1}\sum_{i=1}^n \hat{\xi}_{ui}^2 X_i X_i^T,\ \hat{V}_{n,\varepsilon\xi} = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{ui}\hat{\xi}_{ui} X_i X_i^T. \tag{3.23}$$

The next assumption is required to get strong consistency results of the variance estimators in (3.23).

Assumption 5. *There exist positive constants δ and Δ such that for all i,*

$$E(|X_{ij}^2 X_{ik} X_{il}|^{1+\delta}) < \Delta\ (j, k, l = 1, \ldots, 2 + p + q).$$

Under Assumption 3, 4 and 5, from Theorem 1 of White (1980), it follows that almost surely

$$|\hat{V}_{n,\varepsilon\varepsilon} - \bar{V}_{n,\varepsilon\varepsilon}| \to 0,\ |\hat{V}_{n,\xi\xi} - \bar{V}_{n,\xi\xi}| \to 0,\ |\hat{V}_{n,\varepsilon\xi} - \bar{V}_{n,\varepsilon\xi}| \to 0 \tag{3.24}$$

and that

$$\begin{aligned}
&|(X^TX/n)^{-1}\hat{V}_{n,\varepsilon\varepsilon}(X^TX/n)^{-1} - \bar{M}_n^{-1}\bar{V}_{n,\varepsilon\varepsilon}\bar{M}_n^{-1}| \to 0,\\
&|(X^TX/n)^{-1}\hat{V}_{n,\xi\xi}(X^TX/n)^{-1} - \bar{M}_n^{-1}\bar{V}_{n,\xi\xi}\bar{M}_n^{-1}| \to 0,\\
&|(X^TX/n)^{-1}\hat{V}_{n,\varepsilon\xi}(X^TX/n)^{-1} - \bar{M}_n^{-1}\bar{V}_{n,\varepsilon\xi}\bar{M}_n^{-1}| \to 0,
\end{aligned} \tag{3.25}$$

where the notation is understood to imply convergence of the matrices element by element. Based on (3.25), we can calculate ω_n, $\hat{\alpha}_{yz}^{NC}$ and $\hat{\Sigma}(\hat{\alpha}_{yz}^{NC})$.

3.3 A Test for Instrumental Variable Assumption and Combined Instrumental Variable Estimator

Usual instrumental variable estimators such as two-stage least squares require the models for the outcome and the exposure. One of key assumptions of instrumental variable estimators is that the instrument should be independent of the unmeasured confounder. In contrary, the proposed instrumental variable estimator does not require the model for the exposure, but the model for the negative control outcome, and is still reliable even if the independence assumption is violated.

In this section, first we discuss the method to test the independence assumption by comparing an usual instrumental variable estimator and an instrumental variable estimator with a negative control outcome. Second, under the independence assumption, we introduce an instrumental variable estimator combining both individual instrumental variable estimators. This combined estimator is shown to be more efficient than each instrumental variable estimator.

As an usual instrumental variable estimator, we will consider the two-stage estimator discussed by Choi et al. (2014). Based on the model of (3.3) along with the assumptions made, the model for the main outcome can be written as

$$
\begin{aligned}
Y_i &= \alpha_{yo} + \alpha_{yz} Z_i + \alpha_{yw}^T W_i + \alpha_{yu} U_i + \varepsilon_i \\
&= \alpha_{yo} + \alpha_0^T A_i + \varepsilon_i^*,
\end{aligned} \tag{3.26}
$$

where $\varepsilon_i^* = \alpha_{yu} U_i + \varepsilon_i$, $\alpha_0^T = (\alpha_{yz}, \alpha_{yw}^T)$, $A_i^T = (Z_i, W_i^T)$ and $E(\varepsilon_i \mid A_i, U_i) = 0$.

Assume the following linear model for the exposure:

$$
\begin{aligned}
Z_i &= \beta_{zo} + \beta_{zv}^T V_i + \beta_{zw}^T W_i + \beta_{zu} U_i + \eta_i, \\
&= \beta_{zo} + \beta_0^T D_i + \eta_i^*,
\end{aligned} \tag{3.27}
$$

where $\beta_{zv} = (\beta_{zv,1}, \ldots, \beta_{zv,p})^T$, $\beta_{zw} = (\beta_{zw,1}, \ldots, \beta_{zw,q})^T$, $E(\eta_i \mid D_i, U_i) = 0$ by construction, $\eta_i^* = \beta_{zu} U_i + \eta_i$, $\beta_0^T = (\beta_{zv}^T, \beta_{zw}^T)$, and $D_i^T = (V_i^T, W_i^T)$.

The implied model for A_i is

$$
\begin{aligned}
A_i = \begin{pmatrix} Z_i \\ W_i \end{pmatrix} &= \begin{pmatrix} \beta_{zo} \\ 0_{q\times 1} \end{pmatrix} + \begin{pmatrix} \beta_{zv}^T & \beta_{zw}^T \\ 0_{q\times p} & I_q \end{pmatrix} \begin{pmatrix} V_i \\ W_i \end{pmatrix} + \begin{pmatrix} \eta_i^* \\ 0_{q\times p} \end{pmatrix} \\
&= \beta_{zo}^* + B_0^T D_i + \eta_i^{**},
\end{aligned} \tag{3.28}
$$

where B_0 is a $(p+q) \times (1+q)$ parameter matrix, $0_{q\times p}$ is a $q \times p$ zero matrix and I_q is a $q \times q$ identity matrix.

Substituting model (3.28) into S_i in (3.26) gives

$$
Y_i = \gamma_{yo} + \gamma_0^T D_i + \tau_i, \tag{3.29}
$$

where $\gamma_{yo} = \alpha_{yo} + \alpha_0^T \beta_{zo}^*$ is an intercept, $\gamma_0 = (\gamma_{yv}^T, \gamma_{yw}^T)^T = B_0 \alpha_0$ is a $(p+q) \times 1$ parameter vector and $\tau_i = \varepsilon_i^* + \alpha_0^T \eta_i^{**} = \varepsilon_i^* + \alpha_{yz} \eta_i^*$. The models of (3.27) and (3.29) are called reduced exposure and response models (Choi et al. 2014). The conditional covariance matrix of τ_i and η_i^* given D_i

is allowed to depend on D_i, which includes a special case of a binary exposure (Choi et al. 2014), where the heteroskedastic covariance matrix is induced by a linear probability model error η_i^*.

The two-stage instrumental variable estimator, $\hat{\alpha}^{IV}$, for α_0 is obtained as the form of minimum distance estimator,

$$\hat{\alpha}^{IV} = (\hat{B}^T C_n \hat{B})^{-1} \hat{B}^T C_n \hat{\gamma},$$

where $C_n = C_0 + o(1)$ and $\hat{\theta}^T = (\hat{\gamma}^T, \hat{\beta}^T)$ is a consistent estimator for $\theta_0^T = (\gamma_0^T, \beta_0^T)$. Note that

$$n^{1/2}(\hat{\alpha}^{IV} - \alpha_0) = (\hat{B}^T C_n \hat{B})^{-1} \hat{B}^T C_n n^{\frac{1}{2}} (\hat{\gamma} - \hat{B}\alpha_0).$$

If $\hat{\theta} = \theta_0 + o(1)$ and $n^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically zero-mean normal, then by continuous mapping theorem and Slutsky theorem, $\hat{\alpha}^{IV} = \alpha_0 + o(1)$ and $n^{1/2}(\hat{\alpha}^{IV} - \alpha_0)$ converges in distribution to a zero-mean multivariate normal distribution with the covariance matrix of

$$(B_0^T C_0 B_0)^{-1} B_0^T C_0 \Omega(\alpha_0) C_0 B_0 (B_0^T C_0 B_0)^{-1}, \tag{3.30}$$

where $\Omega(\alpha_0) = \text{var}\left\{n^{1/2}(\hat{\gamma} - \hat{B}\alpha_0)\right\}$. The lower bound of the above covariance matrix of $n^{1/2}(\hat{\alpha}^{IV} - \alpha_0)$ is $(B_0^T \Omega(\alpha_0)^{-1} B_0)^{-1}$, which is obtained by taking $C_0 = \Omega(\alpha_0)^{-1}$. The corresponding $\hat{\alpha}^{IV}$ is obtained by using the weight $C_n = \hat{\Omega}(\hat{\alpha})^{-1}$, which is a consistent estimator for $\Omega(\alpha_0)^{-1}$ if $\hat{\alpha}$ is consistent for α_0. In order to compute $C_n = \hat{\Omega}(\hat{\alpha})^{-1}$, we need an initial consistent estimator for α_0. In practice, we may use the following initial estimator, $\hat{\alpha}_I^{IV} = (\hat{B}^T \hat{B})^{-1} \hat{B}^T \hat{\gamma}$, with identity weight matrix $C_n = I$.

The instrumental variable estimator assumes that $E(\tau_i | D_i) = E(\eta_i^* | D_i) = 0$. This assumption holds if $E(U_i | D_i) = 0$ (Choi et al. 2014). However, if U_i and D_i are correlated, then the condition of $E(U_i | D_i) = 0$ may be violated. Under the model assumptions of (3.1) and (3.2), $\hat{\alpha}_{yz}^{NC}$ is valid whether or not $E(U_i | D_i) = 0$ holds. Thus we can test whether $E(U_i | D_i) = 0$ is satisfied by comparing $\hat{\alpha}_{yz}^{NC}$ and $\hat{\alpha}_{yz}^{IV}$, where $\hat{\alpha}_{yz}^{IV}$ is the first element of $\hat{\alpha}^{IV}$. If $E(U_i | D_i) = 0$ holds, then those two statistics will be similar.

In vector notation, the reduced models can be written as

$$Y = \gamma_{yo} + D\gamma_0^T + \tau \tag{3.31}$$

$$Z = \beta_{zo} + D\beta_0^T + \eta^* \tag{3.32}$$

$Z = (Z_1, ..., Z_n)^T$, $\tau = (\tau_1, ..., \tau_n)^T$, $\eta^* = (\eta_1^*, ..., \eta_n^*)^T$ and D is a design matrix with the ith row of D_i. Without loss of generality, we assume that V_i and W_i are centered to have mean zeros. Then, the least squares estimator for $\theta_0^T = (\gamma_0^T, \beta_0^T)$, $\hat{\theta}^T = (\hat{\gamma}^T, \hat{\beta}^T)$, is given by

$$\hat{\gamma} = (D^T D)^{-1} D^T Y, \tag{3.33}$$

$$\hat{\beta} = (D^T D)^{-1} D^T Z. \tag{3.34}$$

Let D_{ij} denote the jth element of the vector D_i. To obtain a consistency results, we make the following assumption.

Assumption 6. *(a) There exist positive finite constants δ and Δ such that, for all i, $E(|\eta_i^{*2}|^{1+\delta}) < \Delta$; (b) $\bar{L}_n = n^{-1}E(D^T D)$ is nonsingular for (all) n sufficiently large.*

Since $\varepsilon_i^* = \varepsilon_i + \alpha_{yu} U_i$, from Assumption 1(a), it follows that $E(|\varepsilon_i^{*2}|^{1+\delta}) < \Delta$. Again, since $\tau_i = \varepsilon_i^* + \alpha_{yu}\eta_i^*$, from Assumption 1(a) and 4(a), $E(|\tau_i^2|^{1+\delta}) < \Delta$. Assumption 1(a) also ensures that $E(|D_{ij}D_{ik}|^{1+\delta}) < \Delta$ $(j, k = 1, ..., p+q)$. Assumption 1(b) is sufficient for Assumption 4(b).

Theorem 9. *With Assumption 3 and 6, $\hat{\alpha}_{yz}^{IV}$ converges almost surely to α.*

The above theorem is followed by Lemma 1 of White (1980) and Theorem 1 of Choi et al. (2014).

Define the following average covariance matrices

$$\bar{V}_{n,\tau\tau} = n^{-1}\sum_{i=1}^{n} E(\tau_i^2 D_i D_i^T),\ \bar{V}_{n,\eta^*\eta^*} = n^{-1}\sum_{i=1}^{n} E(\eta_i^{*2} D_i D_i^T),\ \bar{V}_{n,\tau\eta^*} = n^{-1}\sum_{i=1}^{n} E(\tau_i \eta_i^* D_i D_i^T),$$

and the covariance matrix

$$\Sigma_\theta = \begin{pmatrix} \bar{L}_n^{-1}\bar{V}_{n,\tau\tau}\bar{L}_n^{-1} & \bar{L}_n^{-1}\bar{V}_{n,\tau\eta^*}\bar{L}_n^{-1} \\ \bar{L}_n^{-1}\bar{V}_{n,\tau\eta^*}\bar{L}_n^{-1} & \bar{L}_n^{-1}\bar{V}_{n,\eta^*\eta^*}\bar{L}_n^{-1} \end{pmatrix}.$$

With the next assumption, an asymptotic normality of $\hat{\alpha}^{IV}$ is obtained.

Assumption 7. *(a) There exist positive finite constants δ and Δ such that for all i $E(|\tau_i^2 D_{ij}D_{ik}|^{1+\delta}) < \Delta$ and $E(|\eta_i^{*2} D_{ij}D_{ik}|^{1+\delta}) < \Delta$ $(j,k = 1,\ldots,p+q)$; (b) Σ_θ is nonsingular for n sufficiently large.*

With Assumption 3, 6 and 7, by Lemma 2 of White (1980), $n^{1/2}(\hat{\theta} - \theta_0)$ converges in distribution to $N(0, \Sigma_\theta)$. Using this result, we obtain the asymptotic normality of the two-stage instrumental variable estimators.

Theorem 10. *Under Assumption 3, 6 and 7, $n^{1/2}(\hat{\alpha}_{yz}^{IV} - \alpha_{yz})$ converges in distribution to a mean-zero multivariate normal distribution with the variance of the $(1,1)$th element of* (3.30).

Define the residuals to be $\hat{\tau} = Y - \hat{\gamma}_{yo} - D\hat{\gamma}_y$ and $\hat{\eta}^* = Z - \hat{\beta}_{zo} - D\hat{\beta}_z$, which are consistent estimators of τ and η^*. The estimators of $V_{n,\tau\tau}$, $V_{n,\eta^*\eta^*}$ and $V_{n,\varepsilon\eta^*}$ are given by

$$\hat{V}_{n,\tau\tau} = n^{-1}\sum_{i=1}^{n}\hat{\tau}_i^2 D_i D_i^T,\ \hat{V}_{n,\eta^*\eta^*} = n^{-1}\sum_{i=1}^{n}\hat{\eta}_i^{*2} D_i D_i^T,\ \hat{V}_{n,\varepsilon\eta^*} = n^{-1}\sum_{i=1}^{n}\hat{\tau}_i\hat{\eta}_i^* D_i D_i^T. \tag{3.35}$$

Under Assumptions 3, 5, 6 and 7(a), from Theorem 1 of White (1980), it follows that almost surely

$$|\hat{V}_{n,\tau\tau} - \bar{V}_{n,\tau\tau}| \to 0,\ |\hat{V}_{n,\eta^*\eta^*} - \bar{V}_{n,\eta^*\eta^*}| \to 0,\ |\hat{V}_{n,\varepsilon\eta^*} - \bar{V}_{n,\varepsilon\eta^*}| \to 0 \tag{3.36}$$

and that

$$|(D^TD/n)^{-1}\hat{V}_{n,\tau\tau}(D^TD/n)^{-1} - \bar{L}_n^{-1}\bar{V}_{n,\tau\tau}\bar{L}_n^{-1}| \to 0,$$
$$|(D^TD/n)^{-1}\hat{V}_{n,\eta^*\eta^*}(D^TD/n)^{-1} - \bar{L}_n^{-1}\bar{V}_{n,\eta^*\eta^*}\bar{L}_n^{-1}| \to 0,$$

$$|(D^T D/n)^{-1}\hat{V}_{n,\tau\eta^*}(D^T D/n)^{-1} - \bar{L}_n^{-1}\bar{V}_{n,\tau\eta^*}\bar{L}_n^{-1}| \to 0, \tag{3.37}$$

where the notation is understood to imply convergence of the matrices element by element.

Under the assumption of $E(U_i|D_i) = 0$, if $n^{1/2}\{(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV}) - (\alpha_{yz}, \alpha_{yz})\}$ converges in distribution to $N\{0, \Sigma(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV})\}$, then we can test whether $E(U_i|D_i) = 0$ using a Wald test statistic. If the absolute Wald test statistic is greater than a certain critical value, say 1.96, then we conclude that $E(U_i|D_i) = 0$ is not true, and this implies the two-stage instrumental variable estimator is not valid. If we do not reject $E(U_i|D_i) = 0$, then we can use both the two-stage instrumental variable estimator and the proposed instrumental variable estimator, and make another instrumental variable estimator by combining those two estimators.

Define the following average covariance matrices:

$$\begin{aligned}
\bar{V}_{n,\varepsilon\tau} &= n^{-1}\sum_{i=1}^{n} E(\varepsilon_{ui}\tau_i X_i D_i^T),\ \bar{V}_{n,\varepsilon\eta^*} = n^{-1}\sum_{i=1}^{n} E(\varepsilon_{ui}\eta_i^* X_i D_i^T),\\
\bar{V}_{n,\xi\tau} &= n^{-1}\sum_{i=1}^{n} E(\xi_{ui}\tau_i X_i D_i^T),\ \bar{V}_{n,\xi\eta^*} = n^{-1}\sum_{i=1}^{n} E(\xi_{ui}\eta_i^* X_i D_i^T).
\end{aligned}$$

Further define

$$\Sigma_{\alpha\theta} = \begin{pmatrix} \bar{M}_n^{-1}\bar{V}_{n,\varepsilon\tau}\bar{L}_n^{-1} & \bar{M}_n^{-1}\bar{V}_{n,\varepsilon\eta^*}\bar{L}_n^{-1} \\ \bar{M}_n^{-1}\bar{V}_{n,\xi\eta^*}\bar{L}_n^{-1} & \bar{M}_n^{-1}\bar{V}_{n,\xi\eta^*}\bar{L}_n^{-1} \end{pmatrix},\ \Sigma_J = \begin{pmatrix} \Sigma_\alpha & \Sigma_{\alpha\theta} \\ \Sigma_{\alpha\theta}^T & \Sigma_\theta \end{pmatrix}. \tag{3.38}$$

By Lemma 2 of White (1980), $n^{1/2}\{(\hat{\alpha}^T, \hat{\theta}^T)^T - (\alpha^{*T}, \theta_0^T)^T)\}$ converges to a mean-zero multivariate normal distribution with the covariance matrix Σ_J under the following assumption together with Assumptions 3, 4, 6, 7.

Assumption 8. *The covariance matrix Σ_J is nonsingular for n sufficiently large.*

The estimators of $\bar{V}_{n,\varepsilon\tau}$, $\bar{V}_{n,\varepsilon\eta^*}$, $\bar{V}_{n,\xi\tau}$ and $\bar{V}_{n,\xi\eta^*}$ are given by

$$\hat{V}_{n,\varepsilon\tau} = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{ui}\hat{\tau}_i X_i D_i^T,\ \hat{V}_{n,\varepsilon\eta^*} = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{ui}\hat{\eta}_i^* X_i D_i^T$$

$$\hat{V}_{n,\xi\tau} = n^{-1}\sum_{i=1}^{n}\hat{\xi}_{ui}\hat{\tau}_i X_i D_i^T,\ \hat{V}_{n,\xi\eta^*} = n^{-1}\sum_{i=1}^{n}\hat{\xi}_{ui}\hat{\eta}_i^* X_i D_i^T$$

Under Assumptions 3, 4(a), 5, 6 and 7(a), from Theorem 1 of White (1980), it follows that almost surely

$$|\hat{V}_{n,\varepsilon\tau} - \bar{V}_{n,\varepsilon\tau}| \to 0,\ |\hat{V}_{n,\varepsilon\eta^*} - \bar{V}_{n,\varepsilon\eta^*}| \to 0,\ |\hat{V}_{n,\xi\tau} - \bar{V}_{n,\xi\tau}| \to 0,\ |\hat{V}_{n,\xi\eta^*} - \bar{V}_{n,\xi\eta^*}| \to 0 \tag{3.39}$$

and that

$$\begin{aligned}
|(X^TX/n)^{-1}\hat{V}_{n,\varepsilon\tau}(D^TD/n)^{-1} - \bar{M}_n^{-1}\bar{V}_{\varepsilon\tau}\bar{L}_n^{-1}| &\to 0,\\
|(X^TX/n)^{-1}\hat{V}_{n,\varepsilon\eta^*}(D^TD/n)^{-1} - \bar{M}_n^{-1}\bar{V}_{\varepsilon\eta^*}\bar{L}_n^{-1}| &\to 0,\\
|(X^TX/n)^{-1}\hat{V}_{n,\xi\tau}(D^TD/n)^{-1} - \bar{M}_n^{-1}\bar{V}_{\xi\tau}\bar{L}_n^{-1}| &\to 0,\\
|(X^TX/n)^{-1}\hat{V}_{n,\xi\eta^*}(D^TD/n)^{-1} - \bar{M}_n^{-1}\bar{V}_{n,\xi\eta^*}\bar{L}_n^{-1}| &\to 0.
\end{aligned} \tag{3.40}$$

Define the $2 \times 4(1+p+q)$ matrix ϕ to be

$$\phi(\alpha^*, \theta_0)^T = \begin{pmatrix} k(\alpha^*)^T & 0_{2(p+q)}^T \\ 0_{2(2+p+q)}^T & f(\theta_0)^T \end{pmatrix}, \tag{3.41}$$

where 0_p is a $p \times 1$ vector of zeros and

$$\begin{aligned}
k(\alpha^*)^T &= (1,\ -\omega_0^T)\dot{g}(\alpha^*)^T\\
f(\theta_0)^T &= \begin{pmatrix}1, & 0_q^T\end{pmatrix}(B_0^TC_0B_0)^{-1}B_0^TC_0\begin{pmatrix}I_{p+q}, & -\alpha_{yz}I_{p+q}\end{pmatrix}.
\end{aligned}$$

Then, $\Sigma(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV})$ is given by $\phi(\alpha^*, \theta_0)^T\Sigma_J\phi(\alpha^*, \theta_0)$. If we have one instrumental variable $(p = 1)$ and B_0 is invertible, then $\hat{\alpha}^{IV} = \hat{B}^{-1}\hat{\gamma}$ dose not depend on the weight matrix C_n. In this case, $f(\theta_0)^T$ has the simple form as $(1,\ 0_q^T)B_0^{-1}\begin{pmatrix}I_{1+q}, & -\alpha_{yz}I_{1+q}\end{pmatrix}$. The Wald test statistic to test

$E(U_i|D_i) = 0$ is

$$\frac{\hat{\alpha}_{yz}^{NC} - \hat{\alpha}_{yz}^{IV}}{\hat{\Sigma}(\hat{\alpha}_{yz}^{NC} - \hat{\alpha}_{yz}^{IV})^{\frac{1}{2}}},$$

where $\hat{\Sigma}(\hat{\alpha}_{yz}^{NC} - \hat{\alpha}_{yz}^{IV}) = (1,-1)\hat{\Sigma}(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV})(1,-1)^T$ and

$$\begin{aligned}
\hat{\Sigma}(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV}) &= \phi(\hat{\alpha}, \hat{\theta})^T \hat{\Sigma}_J \phi(\hat{\alpha}, \hat{\theta}), \\
k(\hat{\alpha})^T &= (1, \ -\omega_n^T)\dot{g}(\hat{\alpha})^T \\
f(\hat{\theta})^T &= \begin{pmatrix} 1, & 0_q^T \end{pmatrix} (\hat{B}^T C_n \hat{B})^{-1} \hat{B}^T C_n \begin{pmatrix} I_{p+q}, & -\hat{\alpha}_{yz}^{IV}(I) I_{p+q} \end{pmatrix}.
\end{aligned}$$

and $\hat{\alpha}_{yz}^{IV}(I)$ is the first element of $\hat{\alpha}_I^{IV}$.

Now we introduce a new instrumental variable estimator combining $\hat{\alpha}_{yz}^{NC}$ and $\hat{\alpha}_{yz}^{IV}$ with the assumption of $E(U_i|D_i) = 0$. We consider a weighted sum of $\hat{\alpha}_{yz}^{NC}$ and $\hat{\alpha}_{yz}^{IV}$. It can be easily shown that the sum of weights should be one to make the combined estimator be consistent. Thus, our new estimator has the form of

$$\hat{\alpha}_{yz}(\lambda_n) = \lambda_n \hat{\alpha}_{yz}^{NC} + (1 - \lambda_n)\hat{\alpha}_{yz}^{IV}, \tag{3.42}$$

where $\lambda_n = \lambda_0 + o(1)$ may depend on the data. Define the estimator with the known weight λ_0 as

$$\hat{\alpha}_{yz}(\lambda_0) = \lambda_0 \hat{\alpha}_{yz}^{NC} + (1 - \lambda_0)\hat{\alpha}_{yz}^{IV}. \tag{3.43}$$

Note that

$$n^{\frac{1}{2}}\{\hat{\alpha}_{yz}(\lambda_n) - \alpha_{yz}\} = n^{\frac{1}{2}}\{\hat{\alpha}_{yz}(\lambda_n) - \hat{\alpha}_{yz}(\lambda_0)\} + n^{\frac{1}{2}}\{\hat{\alpha}_{yz}(\lambda_0) - \alpha_{yz}\} \tag{3.44}$$

and that $n^{1/2}\{\hat{\alpha}_{yz}(\lambda_n) - \hat{\alpha}_{yz}(\lambda_0)\} = n^{1/2}(\hat{\alpha}_{yz}^{NC} - \hat{\alpha}_{yz}^{IV})(\lambda_n - \lambda_0) = o_p(1)$. Thus the distribution of $n^{1/2}\{\hat{\alpha}_{yz}(\lambda_n) - \alpha_{yz}\}$ converges almost surely to the asymptotic distribution of $n^{1/2}\{\hat{\alpha}_{yz}(\lambda_0) -$

$\alpha_{yz}\}$.

The asymptotic covariance matrix of $n^{1/2}\{\hat{\alpha}_{yz}(\lambda_0) - \alpha_{yz}\}$ is

$$\Sigma_\lambda = \lambda_0^2 \Sigma(\hat{\alpha}_{yz}^{NC}) + (1-\lambda_0)^2 \Sigma(\hat{\alpha}_{yz}^{IV}) + 2\lambda_0(1-\lambda_0)\Sigma(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV}). \tag{3.45}$$

The value of λ_0 to yield minimum of Σ_λ is given by

$$\lambda_0 = \frac{\Sigma(\hat{\alpha}_{yz}^{IV}) - \Sigma(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV})}{\Sigma(\hat{\alpha}_{yz}^{NC}) + \Sigma(\hat{\alpha}_{yz}^{IV}) - 2\Sigma(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV})}. \tag{3.46}$$

The value of Σ_λ with (3.46) is given by

$$\Sigma_{\lambda_0} = \Sigma(\hat{\alpha}_{yz}^{IV}) - \frac{\{\Sigma(\hat{\alpha}_{yz}^{IV}) - \Sigma(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV})\}^2}{\Sigma(\hat{\alpha}_{yz}^{NC} - \hat{\alpha}_{yz}^{IV})} = \frac{\Sigma(\hat{\alpha}_{yz}^{NC})\Sigma(\hat{\alpha}_{yz}^{IV}) - \Sigma(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV})^2}{\Sigma(\hat{\alpha}_{yz}^{NC} - \hat{\alpha}_{yz}^{IV})}. \tag{3.47}$$

The equation (3.47) implies that Σ_{λ_0} is smaller than both $\Sigma(\hat{\alpha}_{yz}^{NC})$ and $\Sigma(\hat{\alpha}_{yz}^{IV})$ and that $\Sigma_{\lambda_0} > 0$. We choose λ_n to be consistent estimator of λ_0, namely,

$$\lambda_n = \frac{\hat{\Sigma}(\hat{\alpha}_{yz}^{IV}) - \hat{\Sigma}(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV})}{\hat{\Sigma}(\hat{\alpha}_{yz}^{NC}) + \hat{\Sigma}(\hat{\alpha}_{yz}^{IV}) - 2\hat{\Sigma}(\hat{\alpha}_{yz}^{NC}, \hat{\alpha}_{yz}^{IV})}. \tag{3.48}$$

Following theorems are followed by previous theorems.

Theorem 11. *With Assumptions 3 and 6, and $\lambda_n = \lambda_0 + o(1)$, $\hat{\alpha}_{yz}(\lambda_n)$ converges almost surely to α_{yz}.*

Theorem 12. *With Assumptions 3, 4, 6, 7 and 8, $n^{1/2}\{\hat{\alpha}_{yz}(\lambda_n) - \alpha_{yz}\}$ converges in distribution to $N(0, \Sigma_{\lambda_0})$.*

3.4 Simulations

In the first simulation study, we compare the proposed equally weighted and optimally weighted instrumental variable estimators of the exposure effect. For this comparison, we consider the case where we have multiple instruments. The outcomes and the exposure were generated by the following models.

$$Y_i = \alpha_{yz}Z_i + \alpha_{yv}^T V_i + \alpha_{yw}W_i + \alpha_{yu}U_i + \varepsilon_i, \tag{3.49}$$

$$N_i = \alpha_{nz}Z_i + \alpha_{nv}^T V_i + \alpha_{nw}W_i + \alpha_{nu}U_i + \xi_i, \tag{3.50}$$

$$Z_i = \beta_{zv}^T V_i + \beta_{zw}W_i + \beta_{zu}U_i + \eta_i. \tag{3.51}$$

All of the paraemters are set to be ones except that $\alpha_{nz} = \alpha_{yv} = \alpha_{nv} = 0$ and $V_i = (V_{i1}, V_{i2})^T$ is a two-dimensional vector. The covariate vector $(V_{i1}, V_{i2}, W_i, U_i)$ was generated from standard normal distribution truncated at ± 2.5 and the error vector $(\varepsilon_i, \xi_i, \eta_i)$ was generated from standard normal. We also considered the case of binary exposure by replacing Z_i in (3.49) and (3.50) with $I(Z_i > 0)$. We varied the value of β_{zv} to control the degree of the balance of instruments. For the case of continuous exposure, we used $\beta_{zv} = (1, 1)^T$ to generate balanced instruments and $\beta_{zv} = (0.75, 1.25)^T$ to generate unbalanced instruments. For the case of binary exposure, we used $\beta_{zv} = (2, 2)^T$ to generate balanced instruments and $\beta_{zv} = (1.5, 2.5)^T$ to generate unbalanced instruments.

We summaried the results for the continous exposure in Table 5.1 and those for the binary exposure in Table 3.2. Both tables show that the proposed instruemtnal variable estimator is unbiased and has resonable accuracy in estimating standard errors and confidnece intervals. In the case of contionus exposure, equall and optimal weiths performed similarly. However, in that of binary exposure, the optimal weights outperformed the equall weights

In the second simulation study, we evaluate the combined instrumental variable estimator. The outcome models of the first simulation have been slightly changed to be

$$Y_i = \alpha_{yz}Z_i + \alpha_{yv}V_i + \alpha_{yw}W_i + (\alpha_u + 0.1)U_i + \varepsilon_i, \tag{3.52}$$

$$N_i = \alpha_{nz}Z_i + \alpha_{nv}V_i + \alpha_{nw}W_i + (\alpha_u - 0.1)U_i + \xi_i, \tag{3.53}$$

where V_i is now univariate and the regression parameters for U_i are different between the two outcome models. All of the parameters in (3.52) and (3.53) are set be ones except that $\alpha_{nz} = \alpha_{yv} = \alpha_{nv} = 0$. We varied the value of $\alpha_u \in \{0.5, 1, 2\}$ to control the strength of unmeasured confounding. The continuous exposure was generated from (3.51) with all of the parameters being ones. The covariate vector (V_i, W_i, U_i) was generated from standard normal distribution truncated at ± 2.5 and the error vector $(\varepsilon_i, \xi, \eta_i)$ was generated from standard normal. The binary exposure was generated from Beroulli(P_i), where $P_i = \beta_{zv}^T V_i + \beta_{zw} W_i + \beta_{zu} U_i$ and $(\beta_{zv}, \beta_{zw}, \beta_{zu}) = (0.4, 0.2, 0.2)$, with $(V_i, W_i, U_i) \sim$ Beroulli(0.5) and $(\varepsilon_i, \xi_i) \sim N(0, 0.3^2)$. To estimate α_{yz}, we computed the proposed instrumental variable estimate using a negative control outcome, two-stage instrumental variable estimate and combined instrumental variable estimate.

We summarized the results in Table 3.3.

In the last simulation study, we evaluate the proposed method to test the instrumental variable assumption, $E(U_i|D_i) = 0$. The outcomes and continuous exposure were generated from the same models used for the first simulation except that V_i was univariate and the covariate vector (V_i, W_i, U_i) were generated from truncated zero-mean multivariate normal distribution with the covariance matrix of

$$\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}, \tag{3.54}$$

where $\rho \in \{0, 0.1, 0.2, 0.4\}$ and lower and upper truncation points were $(-2.5, 2.5)$. We computed the proposed instrumental variable estimate and the two-stage instrumental variable estimate and tested whether $E(U_i|D_i) = 0$ is true with the Wald test statistic.

The results for the test of the instrumental variable assumption are summarized in Table 3.4.

Table 3.1: Results for continuous exposure: empirical bias (Bias), empirical standard error (ESE), average of the estimated standard error (ASE) and empirical coverage rate (ECR) of 95% Wald-type confidence interval.

		balaned instruments				unbalaned instruments			
n	Methods	Bias	ESE	ASE	ECR	Bias	ESE	ASE	ECR
400	Equal	-0.011	0.056	0.054	0.936	-0.013	0.065	0.054	0.912
	Optimal	0.000	0.057	0.052	0.928	0.005	0.055	0.058	0.950
	Naive	0.243	0.034	0.034	0.000	0.236	0.033	0.033	0.000
800	Equal	-0.004	0.039	0.038	0.930	-0.004	0.043	0.038	0.910
	Optimal	0.001	0.040	0.036	0.922	0.003	0.040	0.040	0.948
	Naive	0.242	0.024	0.024	0.000	0.235	0.024	0.023	0.000
1600	Equal	-0.001	0.027	0.027	0.952	-0.002	0.029	0.027	0.928
	Optimal	0.001	0.027	0.025	0.936	0.002	0.027	0.028	0.960
	Naive	0.244	0.017	0.017	0.000	0.237	0.017	0.017	0.000
3200	Equal	0.000	0.018	0.019	0.964	0.000	0.020	0.019	0.934
	Optimal	0.001	0.019	0.018	0.942	0.002	0.019	0.020	0.962
	Naive	0.244	0.012	0.012	0.000	0.237	0.011	0.012	0.000
6400	Equal	-0.001	0.014	0.013	0.956	-0.001	0.015	0.014	0.920
	Optimal	0.000	0.014	0.013	0.926	0.000	0.014	0.014	0.954
	Naive	0.244	0.008	0.008	0.000	0.237	0.008	0.008	0.000

Note: “Equal" is the proposed instrumental variable estimator with equal weights, $\omega_n = (0.5, 0.5)^T$. “Optimal" is the proposed instrumental variable estimator with optimal weights. “Naive" is the least squares estimator with observed variables.

Table 3.2: Results for binary exposure: empirical bias (Bias), empirical standard error (ESE), average of the estimated standard error (ASE) and empirical coverage rate (ECR) of 95% Wald-type confidence interval.

		balaned instruments				unbalaned instruments			
n	Methods	Bias	ESE	ASE	ECR	Bias	ESE	ASE	ECR
400	Equal	-0.115	0.642	0.235	0.910	-0.115	0.771	0.241	0.872
	Optimal	0.044	0.231	0.234	0.966	0.045	0.223	0.213	0.942
	Naive	0.476	0.139	0.140	0.078	0.463	0.139	0.140	0.088
800	Equal	-0.028	0.212	0.167	0.938	-0.044	0.228	0.171	0.906
	Optimal	0.023	0.156	0.163	0.946	0.024	0.155	0.147	0.934
	Naive	0.473	0.096	0.099	0.002	0.460	0.096	0.099	0.004
1600	Equal	-0.010	0.109	0.118	0.966	-0.013	0.126	0.121	0.940
	Optimal	0.015	0.104	0.113	0.964	0.017	0.103	0.102	0.956
	Naive	0.481	0.073	0.070	0.000	0.467	0.075	0.070	0.000
3200	Equal	-0.004	0.073	0.083	0.976	-0.005	0.086	0.086	0.950
	Optimal	0.009	0.071	0.079	0.974	0.009	0.072	0.071	0.950
	Naive	0.479	0.050	0.050	0.000	0.468	0.052	0.049	0.000
6400	Equal	-0.005	0.055	0.059	0.966	-0.007	0.062	0.061	0.938
	Optimal	0.001	0.055	0.056	0.956	0.003	0.055	0.050	0.924
	Naive	0.482	0.034	0.035	0.000	0.468	0.034	0.035	0.000

Note: “Equal" is the proposed instrumental variable estimator with equal weights, $\omega_n = (0.5, 0.5)^T$. “Optimal" is the proposed instrumental variable estimator with optimal weights. “Naive" is the least squares estimator with observed variables.

Table 3.3: Results for the combined estimator: strength of unmeasured confounding (Strength), empirical bias (Bias), empirical standard error (ESE), average of the estimated standard error (ASE) and empirical coverage rate (ECR) of 95% Wald-type confidence interval.

			continuous exposure				binary exposure			
Strength	n	Methods	Bias	ESE	ASE	ECR	Bias	ESE	ASE	ECR
α_u = 1	800	iv.nc	-0.004	0.063	0.060	0.954	-0.018	0.188	0.233	0.972
(moderate)		iv.ts	-0.004	0.055	0.054	0.952	0.004	0.113	0.112	0.952
		iv.cm	0.004	0.049	0.047	0.940	0.033	0.085	0.075	0.884
		naive	0.353	0.027	0.028	0.000	0.241	0.044	0.045	0.000
	1600	iv.nc	-0.002	0.043	0.042	0.940	-0.013	0.105	0.082	0.960
		iv.ts	0.000	0.039	0.038	0.958	-0.005	0.078	0.079	0.946
		iv.cm	0.003	0.034	0.033	0.964	0.014	0.055	0.053	0.934
		naive	0.356	0.020	0.020	0.000	0.239	0.033	0.032	0.000
	3200	iv.nc	-0.001	0.029	0.029	0.954	-0.001	0.047	0.045	0.946
		iv.ts	-0.001	0.026	0.027	0.958	0.000	0.057	0.055	0.940
		iv.cm	0.001	0.023	0.024	0.948	0.009	0.040	0.038	0.930
		naive	0.354	0.013	0.014	0.000	0.238	0.022	0.023	0.000
α_u = 2	800	iv.nc	-0.001	0.058	0.056	0.948	-0.002	0.178	0.192	0.974
(strong)		iv.ts	-0.007	0.083	0.083	0.952	0.002	0.198	0.194	0.942
		iv.cm	0.003	0.054	0.051	0.942	0.033	0.098	0.083	0.918
		naive	0.674	0.039	0.041	0.000	0.457	0.079	0.078	0.000
	1600	iv.nc	-0.001	0.040	0.039	0.946	-0.005	0.069	0.065	0.976
		iv.ts	-0.001	0.058	0.059	0.950	-0.007	0.138	0.137	0.950
		iv.cm	0.002	0.036	0.036	0.956	0.012	0.058	0.057	0.946
		naive	0.679	0.028	0.029	0.000	0.458	0.053	0.055	0.000
	3200	iv.nc	0.000	0.027	0.028	0.952	0.001	0.043	0.042	0.950
		iv.ts	-0.001	0.039	0.042	0.962	-0.001	0.100	0.097	0.946
		iv.cm	0.001	0.025	0.026	0.952	0.007	0.040	0.039	0.948
		naive	0.677	0.019	0.020	0.000	0.457	0.041	0.039	0.000

“iv.nc" is the IV estimator with a negative control outcome. “iv.ts" is the two-stage IV estimator. “iv.cm" is the combined IV estimator. “naive" is the least squares estimator with observed variables.

Table 3.4: Results for the test: power of the test (Power), empirical bias (Bias), empirical standard error (ESE), average of the estimated standard error (ASE) and empirical coverage rate (ECR) of 95% Wald-type confidence interval.

ρ	n	Power	Method	Bias	ESE	ASE	ECR
0	800	0.040	iv.nc	-0.002	0.055	0.054	0.942
			iv.ts	-0.001	0.053	0.052	0.944
			naive	0.322	0.027	0.027	0.000
	1600	0.050	iv.nc	0.001	0.036	0.037	0.958
			iv.ts	0.000	0.038	0.036	0.934
			naive	0.323	0.020	0.019	0.000
	3200	0.060	iv.nc	0.001	0.026	0.026	0.946
			iv.ts	0.000	0.026	0.026	0.960
			naive	0.323	0.013	0.013	0.000
	6400	0.052	iv.nc	-0.001	0.018	0.019	0.954
			iv.ts	0.000	0.018	0.018	0.946
			naive	0.323	0.009	0.009	0.000
0.2	800	0.750	iv.nc	-0.006	0.074	0.070	0.934
			iv.ts	0.131	0.041	0.043	0.146
			naive	0.334	0.024	0.025	0.000
	1600	0.966	iv.nc	-0.002	0.047	0.048	0.960
			iv.ts	0.129	0.029	0.030	0.006
			naive	0.334	0.017	0.018	0.000
	3200	1.000	iv.nc	-0.002	0.032	0.034	0.964
			iv.ts	0.132	0.022	0.021	0.000
			naive	0.334	0.013	0.013	0.000
	6400	1.000	iv.nc	0.000	0.023	0.024	0.956
			iv.ts	0.132	0.015	0.015	0.000
			naive	0.335	0.009	0.009	0.000
0.4	800	0.894	iv.nc	-0.023	0.138	0.120	0.960
			iv.ts	0.208	0.039	0.039	0.000
			naive	0.332	0.025	0.025	0.000
	1600	1.000	iv.nc	-0.015	0.073	0.075	0.964
			iv.ts	0.206	0.027	0.028	0.000
			naive	0.330	0.017	0.018	0.000
	3200	1.000	iv.nc	-0.008	0.052	0.051	0.966
			iv.ts	0.207	0.020	0.019	0.000
			naive	0.330	0.012	0.012	0.000
	6400	1.000	iv.nc	-0.001	0.034	0.035	0.952
			iv.ts	0.208	0.014	0.014	0.000
			naive	0.331	0.009	0.009	0.000

CHAPTER 4: SENSITIVITY ANALYSIS FOR AN EXPOSURE EFFECT WITH A NEGATIVE CONTROL OUTCOME WHEN UNOBSERVED CONFOUNDERS MAY BE PRESENT

4.1 Introduction

Confounding happens when there are variables affecting both of an outcome and an exposure. To deal with confounding, one can take regression adjustment or propensity score method. However, those methods require that all of the confounding variables are observed and this may not be true in observational studies. A commonly employed method to rule out the bias due to unmeasured confounding is an instrumental variable (IV) method (Brookhart et al. 2006, McClellan et al. 1994, Schneeweiss et al. 2006, Stukel et al. 2007). Although the requirements of an IV depend on a particular analytic method one chooses, we can say that a variable is an IV if it satisfies the following three conditions (Brookhart et al. 2010): (i) it has a causal effect on the exposure; (ii) it has effects on the outcome only through the exposure; (iii) it is unrelated to an unmeasured confounder. A typical example of an IV is a randomized indicator, which is usually available, to estimate a drug effect in randomized trial with non-compliance. However, in observational studies, it is not easy to find an “valid" IV.

In the case where we do not have a tool such a valid IV, thus it is hard do find a reliable way to consistently estimate the parameters, it is desirable to evaluate the sensitivity of regression results to unmeasured confounders. There have been several developed sensitivity analysis techniques (Rosenbaum and Rubin 1983, Lin et al. 1998, Brumback et al. 2004, Gustafson et al. 2010, VanderWeele et al. 2012). Among those, the methods of Lin et al. (1998) are applicable to general regression models and can be easily performed (VanderWeele 2008). Lin et al. (1998) assumed that the distribution of the unmeasured confounder conditional on the measured confounders and the exposure is normal or binomial, and they identified simple algebraic relationships between a true exposure effect in the full model and an apparent exposure effect in the reduced model which does not control for unmeasured confounders. One can make an inference on the true exposure effect by making a simple adjustment to the estimate and the confidence interval of the apparent exposure effect. Lin et al. (1998) developed their method for linear, log-linear, logistic and proportional hazard models.

Negative controls technique has been emerging as a tool to detect unmeasured confounding (Flanders et al. 2011, Lipsitch et al. 2010, Lumley and Sheppard 2000, Jackson et al. 2006, Smith 2008; 2012). An outcome is said to be a valid negative control outcome (N) if it is influenced by measured confounders (W) and an measured confounder (U) in the association between the exposure (Z) and the main outcome (Y), but not directly influenced by the exposure (Lipsitch et al. 2010). Those conditions are sufficient to detect unmeasured confounding, but insufficient to estimate the causal effect. Tchetgen Tchetgen (2014) made more progress to estimate the causal effect by imposing an additional assumption that the negative control outcome is independent of a treatment received conditional on the measured confounders and counterfactual outcomes. This assumption implies that the counterfactual outcomes are ideal proxies of the unmeasured confounders. Under this assumption, an additive causal effect on a continuous outcome can be estimated by regressing N onto (W, Y, Z).

In this paper, we develop new methods of sensitivity analysis in various regression models under unmeasured confounding using a negative control outcome. As in Lin et al. (1998), we assume that the causal exposure effect is represented by a regression parameter. We use outcome models which follow the definition of the negative control outcome presented in Lipsitch et al. (2010) and definition 1 of Tchetgen Tchetgen (2014), but unlike assumption 1 of Tchetgen Tchetgen (2014) the outcomes are not ideal proxies in our model setup, thus the methods of Tchetgen Tchetgen (2014) are not valid. In Section 2, first, we extended the control outcome calibration approach of Tchetgen Tchetgen (2014) to conduct sensitivity analysis. Second, we extended Lin's methods to use a negative control outcome with fewer sensitivity parameters. Furthermore, we developed

Lin's methods for probit model and additive hazard model and extended those to use a negative control outcome. Third, we applied Lin's methods to a conditional likelihood with a negative control outcome for binary and right-censored outcome data. The proposed methods are shown to perform well and better than Lin's method in logistic and proportional hazard regression models reported in Section 3, where the conditional likelihood method is superior than both the marginal model method and the Lin's method when the unmeasured confounding is huge.

4.2 Methods

In this section, we describe the proposed three types of techniques of sensitivity analysis using a negative control outcome. Denote n by the number of samples. For $i = 1, ..., n$, suppose that Y_i is the main outcome, N_i is the negative control outcome, Z_i is the exposure, $W_i = (W_{i1}, ..., W_{iq})^T$ is the $q \times 1$ vector of measured confounders and U_i is the unmeasured confounder.

4.2.1 Control outcome calibration approach

Tchetgen Tchetgen (2014) proposed control calibration approach which fits the model of the negative control outcome conditional on measured variables including the main outcome. This approach can be generalized to conduct a sensitivity analysis by considering the model of $Y_i - cN_i$, where c is a known constant related to a sensitivity parameter.

For the main and negative control outcomes, we consider the following linear models,

$$Y_i = \beta_{yo} + \beta_{yz} Z_i + \beta_{yw}^T W_i + \beta_{yu} U_i + \varepsilon_i = \beta_y^T X_i + \beta_{yu} U_i + \varepsilon_i, \tag{4.1}$$

$$N_i = \beta_{no} + \beta_{nz} Z_i + \beta_{nw}^T W_i + \beta_{nu} U_i + \xi_i = \beta_n^T X_i + \beta_{nu} U_i + \xi_i, \tag{4.2}$$

where for $i = 1, ..., n$, (ε_i, ξ_i) is assumed to be independent and identically distributed error terms with $E(\varepsilon_i | X_i, U_i) = E(\xi_i | X_i, U_i) = 0$, $\beta_{yw} = (\beta_{yw,1}, ..., \beta_{yw,q})^T$ is a $q \times 1$ parameter vector, $\beta_{nz} = 0$, and $X_i^T = (1, Z_i, W_i^T)$. Note that models of Y_i and N_i share the measured and unmeasured confounders and that $\beta_{nz} = 0$, which correspond to the the definition of a negative control outcome described in Lipsitch et al. (2010) and in definition 1 of Tchetgen Tchetgen (2014).

Consider the difference of the outcomes depending on a constant c, $D_i(c) = Y_i - cN_i$. The model of $D_i(c)$ can be written as

$$D_i(c) = (\beta_{yo} - c\beta_{no}) + \beta_{yz}Z_i + (\beta_{yw} - c\beta_{nw})^T W_i + (\beta_{yu} - c\beta_{nu})U_i + (\varepsilon_i - c\xi_i). \tag{4.3}$$

Define the sensitivity parameter: $\gamma_u = \beta_{yu}/\beta_{nu}$, which is the ratio of the regression parameters of U_i. Further define $\gamma_u(c)$ to be

$$\begin{aligned}\gamma_u(c) \equiv \beta_{yu} - c\beta_{nu} &= \beta_{yu} - (c - \gamma_u + \gamma_u)\beta_{nu},\\ &= (\gamma_u - c)\beta_{nu}.\end{aligned} \tag{4.4}$$

Using (4.4), the model (4.3) can be written as

$$\begin{aligned}D_i(c) &= \delta_o(c) + \beta_{yz}Z_i + \delta_w(c)^T W_i + \gamma_u(c)U_i + \tau_i(c),\\ &= \theta(c)^T X_i + \gamma_u(c)U_i + \tau_i(c),\end{aligned} \tag{4.5}$$

where $\delta_o(c) = \beta_{yo} - c\beta_{no}$, $\delta_w(c) = \beta_{yw} - c\beta_{nw}$, $\tau_i(c) = \varepsilon_i - c\xi_i$ and $\theta(c)^T = \{\delta_o(c), \beta_{yz}, \delta_w(c)^T\}$. In vector notation, model (4.5) can be written as

$$D(c) = X\theta(c) + U\gamma_u(c) + \tau(c), \tag{4.6}$$

where $D(c)$, U and $\tau(c)$ are $n \times 1$ vectors with typical rows of $D_i(c)$, U_i and $\tau_i(c)$,and X is a $n \times p$ matrix with a typical row of X_i, where $p = 2 + q$. Then, the least squares estimator of $\hat{\theta}(c)$

is given by

$$\hat{\theta}(c) = (X^T X)^{-1} X^T D(c), \tag{4.7}$$

which is consistent for θ when $c = \gamma_u$. If a prior information about γ_u is available, then we can use it to calculate $\hat{\theta}(c)$. However, that information is not readily available in practice. Conducting a sensitivity analysis may be desirable within a reasonable range of c.

To identify the limiting value of $\hat{\theta}(c)$, we impose the following assumptions.

Assumption 9. *Let $A_i^T = (X_i^T, U_i)$, $\Sigma_n(XX) = n^{-1}E(X^T X)$ and $\Sigma_n(XU) = n^{-1}E(X^T U)$. (a) There exist positive finite constants δ and Δ such that for all i, $E(|\tau_i(c)^2|^{1+\delta}) < \Delta$ and $E(|A_{ij}A_{ik}|^{1+\delta}) < \Delta$ $(j, k = 1, ..., p+1)$; (b) $\Sigma_n(XX)$ is nonsingular for (all) n sufficiently large.*

From Assumption 1(a), it follows by strong law of large numbers that

$$n^{-1}X^T X \rightarrow_{a.s} \Sigma_n(XX),\ n^{-1}X^T U \rightarrow_{a.s} \Sigma_n(XU),\ \text{and}\ n^{-1}X^T \tau(c) \rightarrow_{a.s} 0.$$

Under Assumption 1, the limiting value of $\hat{\theta}(c)$ is given by

$$\hat{\theta}(c) = \theta(c) + \gamma_u(c)(X^T X)^{-1} X^T U + (X^T X)^{-1} X^T \tau(c) \rightarrow_{a.s} \theta_{ols}(c),$$

where $\theta_{ols}(c) = \theta(c) + \gamma_u(c)\omega$ and $\omega = \Sigma_n(XX)^{-1}\Sigma_n(XU)$.

Write

$$n^{\frac{1}{2}}\{\hat{\theta}(c) - \theta_{ols}(c)\} = (n^{-1}X^T X)^{-1} n^{\frac{1}{2}} X^T \tau^*(c),$$

where $\tau^*(c) = Y - X\theta_{ols}(c) = \gamma_u(c)(U - X\omega) + \tau(c)$, which is consistently estimated by $\hat{\tau}^*(c) = Y - X\hat{\theta}(c)$. Note that $E\{X^T \tau^*(c)\} = 0$ while $E\{\tau^*(c)\}$ may not be zero. Thus we use theoretical results of White (1980). With the following assumption, the asymptotic distribution of $\hat{\theta}(c)$ can obtained.

Assumption 10. *(a) There exist positive constant δ and Δ such that for all i, $E(|\tau_i^*(c)^2|^{1+\delta}) < \Delta$*

and $E(|\tau_i^*(c)_i^2 X_{ij} X_{ik}|^{1+\delta}) < \Delta$ $(j,k = 1,...,p)$*; (b) The average covariance matrix* $\bar{V}_n(c) = n^{-1}\sum_{i=1}^n E(\tau_i^*(c)^2 X_i X_i^T)$ *is nonsingular for* n *sufficiently large.*

Under assumption 1 and 2, by Lemma 2 of White (1980), the asymptotic distribution of $n^{\frac{1}{2}}\{\hat{\theta}(c) - \theta_{ols}(c)\}$ is a zero-mean multivariate normal distribution with the covariance matrix of

$$\Sigma_n\{\hat{\theta}(c)\} = \Sigma_n(XX)^{-1}\bar{V}_n(c)\Sigma_n(XX)^{-1}.$$

By Theorem 1(i) of White (1980), the estimators

$$\hat{V}_n(c) = n^{-1}\sum_i^n \hat{\tau}_i^*(c)_i^2 X_i X_i^T, \text{ and } \hat{\Sigma}_n\{\hat{\theta}(c)\} = (n^{-1}X^TX)^{-1}\hat{V}_n(c)(n^{-1}X^TX)^{-1}$$

are consistent for $\bar{V}_n(c)$ and $\Sigma_n\{\hat{\theta}(c)\}$ respectively with the following assumption along with assumption 1 and 2.

Assumption 11. *There exist positive constants* δ *and* Δ *such that for all* i*,*

$$E(|X_{ij}^2 X_{ik} X_{il}|^{1+\delta}) < \Delta \ (j,k,l = 1,...,p).$$

In our model setup, the estimator of Tchetgen Tchetgen (2014) can be obtained by imposing an additional condition on model (4.1), which is related to Assumption 1 of Tchetgen Tchetgen (2014). Note that from (4.1), U_i can be written as

$$U_i = \beta_{yu}^{-1}(Y_i - \beta_y^T X_i - \varepsilon_i). \tag{4.8}$$

Substituting (4.8) into (4.2) gives

$$\begin{aligned} N_i &= \left(\beta_n - \frac{\beta_{nu}}{\beta_{yu}}\beta_y\right)^T X_i + \frac{\beta_{nu}}{\beta_{yu}}Y_i + \xi_i - \frac{\beta_{nu}}{\beta_{yu}}\varepsilon_i, &(4.9)\\ &= \alpha_n^T X_i + \alpha_{ny}Y_i + \zeta_i, &(4.10)\end{aligned}$$

where $\alpha_n = (\alpha_{no}, \alpha_{nz}, \alpha_{nw}^T)^T = \beta_n - (\beta_{nu}/\beta_{yu})\beta_y$, $\alpha_{ny} = \beta_{nu}/\beta_{yu}$ and $\zeta_i = \xi_i - \alpha_{ny}\varepsilon_i$. To make Assumption 1 of Tchetgen Tchetgen (2014) be satisfied, we need the condition that Y_i and ε_i is uncorrelated, which is realized if the variance of ε_i is zero. Under this assumption, the exposure effect estimator is given by

$$-\alpha_{nz}/\alpha_{ny}, \tag{4.11}$$

which is the estimator of Tchetgen Tchetgen (2014). One can see that the inference is made by estimating the sensitivity parameter $\gamma_u = 1/\alpha_{ny}$ directly.

4.2.2 Marginal model approach

Lin et al. (1998) proposed the method to assess the sensitivity of regression results to unmeasured confounders in observational studies. They derived reduced models which are obtained by integrating the full models with respect to specified conditional distribution of the unmeasured confounder given the exposure and the measured confounders. They identified a simple algebraic relationship between the true exposure effect in the full model and the apparent exposure effect in the reduced model.

We extended Lin's approach to the case where a negative control outcome is present. To apply the methods of Lin et al. (1998), we need to assume that the exposure is binary. For continuous outcome, since U_i is unmeasured, one is forced to fit the reduced models,

$$Y_i = \beta_y^{*T} X_i + \varepsilon_i^*, \tag{4.12}$$

$$N_i = \beta_n^{*T} X_i + \xi_i^*, \tag{4.13}$$

where $\beta_y^{*T} = (\beta_{yo}^*, \beta_{yz}^*, \beta_{yw}^{*T})$, $\beta_n^{*T} = (\beta_{no}^*, \beta_{nz}^*, \beta_{nw}^{*T})$ and $E(\varepsilon_i^*|Z_i, W_i) = E(\xi_i^*|Z_i, W_i) = 0$ by construction. Note that

$$E(Y_i|Z_i, W_i) = \beta_y^T X_i + \beta_{yu}\mu(Z_i, W_i), \tag{4.14}$$

where $\mu(Z_i, W_i) = E(U_i|Z_i, W_i)$. Model (4.14) can be written as

$$E(Y_i|Z_i, W_i) = \beta_{yo} + \beta_{yu}\mu(0, W_i) + \{\beta_{yz} + \beta_{yu}\delta(W_i)\}Z_i + \beta_{yw}^T W_i, \tag{4.15}$$

where $\delta(W_i) = \mu(1, W_i) - \mu(0, W_i)$. Thus,

$$\beta_{yz} = \beta_{yz}^* - \beta_{yu}\delta(W_i) \tag{4.16}$$

To make (4.16) more sensible, we assume that the effects of Z_i and W_i on $\mu(Z_i, W_i)$ are additive, i.e., $\mu(Z_i, W_i) = \mu_z(Z_i) + \mu_w(W_i)$ as in Lin et al. (1998), where $\mu_z(\cdot)$ and $\mu_w(\cdot)$ are marginal functionals of Z and W. Then, (4.16) reduces to

$$\beta_{yz} = \beta_{yz}^* - \beta_{yu}\delta, \tag{4.17}$$

where $\delta = \mu_z(1) - \mu_z(0)$. Let $\hat{\beta}_{yz}$ denote the consistent estimator of β_{yz}^*. Define the estimator to be

$$\hat{\beta}_{yz}(\beta_{yu}, \delta) = \hat{\beta}_{yz} - \beta_{yu}\delta. \tag{4.18}$$

Lin et al. (1998) proposed sensitivity analysis using $\hat{\beta}_{yz}(\beta_{yu}, \delta)$ with the sensitivity parameters β_{yu} and δ.

We propose a new method of sensitivity analysis with a negative control outcome. Our method only needs to vary one parameter, $\gamma_u = \beta_{yu}/\beta_{nu}$, which is more intuitive to be varied than β_{yu} and δ. Similarly to (4.16), we can show that

$$\beta_{nz} = \beta_{nz}^* - \beta_{nu}\delta \tag{4.19}$$

Since $\beta_{nz} = 0$ by the definition of the negative control outcome, $\delta = \beta_{nz}^*/\beta_{nu}$. This gives the following formula for β_{yz},

$$\beta_{yz} = \beta^*_{yz} - \gamma_u \beta^*_{nz}. \tag{4.20}$$

We can estimate β^*_{yz} and β^*_{nz} consistently by fitting the models (4.12) and (4.13). Denote $\hat{\beta}_{yz}$ and $\hat{\beta}_{nz}$ by the consistent estimators of β^*_{yz} and β^*_{nz}. Sensitivity analysis is performed by calculating the following estimator with varying γ_u,

$$\hat{\beta}_{yz} - \gamma_u \hat{\beta}_{nz}. \tag{4.21}$$

We can see that (4.21) is equivalent to the control outcome calibration method, which is developed without the requirement that the exposure is binary. In fact, the methods of Lin et al. (1998) can allow continuous Z_i with the additive model assumption of U_i such that $\mu(Z_i, W_i) = \tau Z_i + q(W_i)$ (VanderWeele, 2008). In that case, δ in the formula (4.20) may be replaced by τ.

Lin's approach for continuous outcomes may be more useful when dealing with censoring. Under the accelerated failure time models with right-censoring, we often set Y_i and N_i to be log of survival times. The right-censoring times for Y_i and N_i are assumed to be independent of each outcome conditional on X_i. There are well-established methods for parameter estimation of accelerated failure time model such as rank estimation and Buckley-James estimation. The methods of Jin et al. (2006) for multivariate accelerated failure time models can be applied to estimate the regression parameters and the covariance matrix for models (4.12) and (4.13).

For binary outcomes, we first assume the log-linear models,

$$\text{pr}(Y_i = 1 | Z_i, W_i, U_i) = \exp(\beta_y^T X_i + \beta_{yu} U_i), \tag{4.22}$$

$$\text{pr}(N_i = 1 | Z_i, W_i, U_i) = \exp(\beta_n^T X_i + \beta_{yu} U_i). \tag{4.23}$$

The log-linear model is an good approximation to the logistic regression model for rare events.

Since U_i is unmeasured, one is forced to fit the following reduced models

$$\text{pr}(Y_i = 1|Z_i, W_i) = \exp(\beta_y^{*T} X_i), \tag{4.24}$$

$$\text{pr}(N_i = 1|Z_i, W_i) = \exp(\beta_n^{*T} X_i). \tag{4.25}$$

Let $F(u|Z_i, W_i)$ be the distribution function of U_i given Z_i and W_i. By the law of conditional expectation,

$$\begin{aligned} \text{pr}(Y_i = 1|Z_i, W_i) &= \int_{-\infty}^{\infty} \text{pr}(Y_i = 1|Z_i, W_i, u) dF(u|Z_i, W_i) \\ &= \exp(\beta_y^T X_i) \int_{-\infty}^{\infty} \exp(\beta_{yu} u) dF(u|Z_i, W_i). \end{aligned} \tag{4.26}$$

Similarly,

$$\text{pr}(N_i = 1|Z_i, W_i) = \exp(\beta_n^T X_i) \int_{-\infty}^{\infty} \exp(\beta_{nu} u) dF(u|Z_i, W_i). \tag{4.27}$$

Suppose that conditional on Z_i and W_i, the confounder U_i is normally distributed with mean $\mu(Z_i, W_i)$ and unit variance. Lin et al. (1998) showed that reduced models (4.26) and (4.27) are still log-linear models:

$$\text{pr}(Y_i = 1|Z_i, W_i) = \exp[\beta_{yo} + \beta_{yu}\mu(0, W_i) + 0.5\beta_{yu}^2 + \{\beta_{yz} + \beta_{yu}\delta(W_i)\}Z_i + \beta_{yw}^T W_i], \tag{4.28}$$

$$\text{pr}(N_i = 1|Z_i, W_i) = \exp[\beta_{no} + \beta_{nu}\mu(0, W_i) + 0.5\beta_{nu}^2 + \{\beta_{nz} + \beta_{nu}\delta(W_i)\}Z_i + \beta_{nw}^T W_i], \tag{4.29}$$

where $\delta(W_i) = \mu(1, W_i) - \mu(0, W_i)$. Under the additive model assumption of $\mu(Z_i, W_i)$,

$$\beta_{yz} = \beta_{yz}^* - \beta_{yu}\delta, \tag{4.30}$$

$$\beta_{nz} = \beta_{nz}^* - \beta_{nu}\delta. \tag{4.31}$$

By using that $\beta_{nz} = 0$, a sensitivity analysis can be performed in the same way as in the continuous case. Again, Lin's method needs to vary two parameters β_{yu} and $\delta(W_i)$, but only one parameter γ_u needs to be varied for our method.

Now we consider the logistic regression models for binary outcomes,

$$\text{logit}\{\text{pr}(Y_i = 1|Z_i, W_i, U_i)\} = \beta_y^T X_i + \beta_{yu} U_i, \tag{4.32}$$

$$\text{logit}\{\text{pr}(N_i = 1|Z_i, W_i, U_i)\} = \beta_n^T X_i + \beta_{nu} U_i. \tag{4.33}$$

From models (4.32) and (4.33), the reduced are given as

$$\text{pr}(Y_i|Z_i, W_i) = \exp(\beta_y^T X_i) \int_{-\infty}^{\infty} \frac{\exp(\beta_{yu} u)}{1 + \exp(\beta_y^T X_i + \beta_{yu} u)} dF(u|Z_i, W_i), \tag{4.34}$$

$$\text{pr}(N_i|Z_i, W_i) = \exp(\beta_n^T X_i) \int_{-\infty}^{\infty} \frac{\exp(\beta_{nu} u)}{1 + \exp(\beta_n^T X_i + \beta_{nu} u)} dF(u|Z_i, W_i). \tag{4.35}$$

Generally, equations of (4.34) and (4.35) does not reduce to logistic regression models. In other words, models (4.34) and (4.35) cannot be written as

$$\text{logit}\{\text{pr}(Y_i = 1|Z_i, W_i)\} = \beta_y^{*T} X_i \tag{4.36}$$

$$\text{logit}\{\text{pr}(N_i = 1|Z_i, W_i)\} = \beta_n^{*T} X_i. \tag{4.37}$$

It is practically important to ascertain under what circumstances models (4.36) and (4.37) provide reasonable approximations to models (4.34) and (4.35) and whether simple relationships such as (4.30) and (4.31) exit for the logistic regression. Lin et al. (1998) showed that (4.34) can be written as

$$\text{logit}\{P(Y_i = 1|Z_i, W_i)\} = \beta_y^T X_i + g(Z_i, W_i), \tag{4.38}$$

where

$$g(Z_i, W_i) = \beta_{yu} \mu(Z_i, W_i) + 0.5\beta_{yu}^2$$

$$+\log\frac{\int_{-\infty}^{\infty}\{1+\exp(\beta_{yu}^2+\beta_y^TX_i+\beta_{yu}u)\}^{-1}\exp\{-\frac{u-\mu(Z_i,W_i)^2}{2}\}du}{\int_{-\infty}^{\infty}\{1+\exp(\beta_y^TX_i+\beta_{yu}u)\}^{-1}\exp\{-\frac{u-\mu(Z_i,W_i)^2}{2}\}du} \tag{4.39}$$

If the event is rare or $|\beta_{yu}|$ is small, $g(Z_i,W_i)\approx\beta_{yu}\mu(Z_i,W_i)+0.5\beta_{yu}^2$. It then follows that

$$\text{logit}\{P(Y_i=1|Z_i,W_i)\}\approx\beta_{yo}+\beta_{yu}\mu(0,W_i)+0.5\beta_{yu}^2+\{\beta_{yz}+\beta_{yu}\delta(W_i)\}Z_i+\beta_{yw}^TW_i \tag{4.40}$$

Model (4.40) takes the form of (4.36) and imply that formulas (4.30) and (4.31) hold approximately for the logistic regression. Then both formulas (4.18) and (4.20) for sensitivity analysis also approximately hold.

An alternative to the logistic regression model is the probit regression model. Our probit regression models for the main and negative control outcomes are given by

$$\text{pr}(Y_i=1|Z_i,W_i,U_i)=\Phi(\beta_y^TX_i+\beta_{yu}U_i), \tag{4.41}$$

$$\text{pr}(N_i=1|Z_i,W_i,U_i)=\Phi(\beta_n^TX_i+\beta_{nu}U_i), \tag{4.42}$$

where $\Phi(\cdot)$ is a cumulative normal distribution function. One is forced to fit

$$\text{pr}(Y_i=1|Z_i,W_i)=\Phi(\beta_y^{*T}X_i), \tag{4.43}$$

$$\text{pr}(N_i=1|Z_i,W_i)=\Phi(\beta_n^{*T}X_i). \tag{4.44}$$

Using the result of Carroll and Spiegelman (1984), we can obtain the explicit form of $\text{pr}(Y_i=1|Z_i,W_i)$:

$$\begin{aligned}\text{pr}(Y_i=1|Z_i,W_i)&=\Phi\left\{\frac{\beta_{yo}+\beta_{yu}\mu(Z_i,W_i)+\beta_{yz}Z_i+\beta_{yw}^TW_i}{(1+\beta_{yu}^2)^{\frac{1}{2}}}\right\}\\&=\Phi\left[\frac{\beta_{yo}+\beta_{yu}\mu(0,W_i)+\{\beta_{yz}+\beta_{yu}\delta(W_i)\}Z_i+\beta_{yw}^TW_i}{(1+\beta_{yu}^2)^{\frac{1}{2}}}\right].\end{aligned} \tag{4.45}$$

Under the additive model assumption of $\mu(Z_i, W_i)$, it follows that

$$\beta_{yz} = (1 + \beta_{yu}^2)^{\frac{1}{2}} \beta_{yz}^* - \beta_{yu}\delta \tag{4.46}$$

Unfortunately, using a negative control outcome does not reduce the number of the sensitivity parameters. To see this, note that

$$\beta_{nz} = (1 + \beta_{nu}^2)^{\frac{1}{2}} \beta_{nz}^* - \beta_{nu}\delta, \tag{4.47}$$

which implies $\delta = \beta_{nu}^{-1}(1 + \beta_{nu}^2)^{1/2}\beta_{nz}^*$. Substitution gives that

$$\beta_{yz} = (1 + \beta_{yu}^2)^{\frac{1}{2}} \beta_{yz}^* - (\gamma_u^2 + \beta_{yu}^2)^{\frac{1}{2}} \beta_{nz}^*, \tag{4.48}$$

which has two sensitivity parameters, γ_u and β_{yu}.

Now we discuss the method for survival time outcomes which are subject to be right-censored. Let T_{yi} and T_{ni} denote the main and negative control survival times for the ith subject. The true proportional hazard models conditional on (Z_i, W_i, U_i) are given by

$$\lambda_y(t|Z_i, W_i, U_i) = \lambda_0^Y(t) \exp(\beta_y^T X_i + \beta_{yu} U_i), \tag{4.49}$$

$$\lambda_n(t|Z_i, W_i, U_i) = \lambda_0^N(t) \exp(\beta_n^T X_i + \beta_{nu} U_i), \tag{4.50}$$

where X_i does not include an intercept and $\lambda_0^Y(t)$ and $\lambda_0^N(t)$ are arbitrary baseline hazard functions. Since U_i is unmeasured, one is forced to fit

$$\lambda_y(t|Z_i, W_i) = \lambda_0^{Y*}(t) \exp(\beta_y^{*T} X_i), \tag{4.51}$$

$$\lambda_n(t|Z_i, W_i) = \lambda_0^{N*}(t) \exp(\beta_n^{*T} X_i) \tag{4.52}$$

where $\lambda_0^{Y*}(t)$ and $\lambda_0^{N*}(t)$ are arbitrary baseline hazard functions. Let $f(t|\cdot)$ and $S(t|\cdot)$ be the conditional density and survival functions of T_{yi}. By elementary probability arguments,

$$\lambda_y(t|Z_i, W_i) = \frac{f(t|Z_i, W_i)}{S(t|Z_i, W_i)} = \frac{\int_{-\infty}^{\infty} f(t|Z_i, W_i, u) dF(u|Z_i, W_i)}{\int_{-\infty}^{\infty} S(t|Z_i, W_i, u) dF(u|Z_i, W_i)} \tag{4.53}$$

Lin et al. (1998) showed that

$$\lambda_y(t|Z_i, W_i) = \lambda_0^Y(t) \exp(\beta_y^T X_i) h(t; Z_i, W_i), \tag{4.54}$$

where

$$\begin{aligned} h(t; Z_i, W_i) = &\exp\{\beta_{yu}\mu(Z_i, W_i) + 0.5\beta_{yu}^2\} \\ &\frac{\int_{-\infty}^{\infty} \exp\{-\Lambda_0^Y(t) \exp(\beta_{yu}^2 + \beta_y^T X_i + \beta_{yu} u)\} \exp\{-\frac{u-\mu(Z_i,W_i)^2}{2}\} du}{\int_{-\infty}^{\infty} \exp\{-\Lambda_0^Y(t) \exp(\beta_y^T X_i + \beta_{yu} u)\} \exp\{-\frac{u-\mu(Z_i,W_i)^2}{2}\} du}, \end{aligned} \tag{4.55}$$

where $\Lambda_0^Y(t) = \int_0^t \lambda_0^Y(t)$. It can be approximated by $h(t; Z_i, W_i) \approx \exp\{\beta_{yu}\mu(Z_i, W_i) + 0.5\beta_{yu}^2\}$ if $\Lambda_0(t)$ or $|\beta_{yu}|$ is small. Then

$$\lambda_y(t|Z_i, W_i) \approx \lambda_0^Y(t) \exp[\beta_{yu}\mu(0, W_i) + \{\beta_{yz} + \beta_{yu}\delta(W_i)\} Z_i + \beta_{yw}^T W_i], \tag{4.56}$$

which has the form of (4.51). Then both formulas (4.18) and (4.20) for sensitivity analysis also approximately hold.

An alternative to the proportional hazard model is the additive hazard model. As shown in Lin et al. (1998), generally, marginalization does not keep the original model under the proportional hazard model. However, we will show that this is the case under additive hazard model.

The true additive hazard models conditional on (Z_i, W_i, U_i) are given by

$$\lambda_y(t|Z_i, W_i, U_i) = \lambda_0^Y(t) + \beta_y^T X_i + \beta_{yu} U_i, \tag{4.57}$$

$$\lambda_n(t|Z_i, W_i, U_i) = \lambda_0^N(t) + \beta_n^T X_i + \beta_{nu} U_i, \tag{4.58}$$

where X_i does not include an intercept and $\lambda_0^Y(t)$ and $\lambda_0^N(t)$ are arbitrary baseline hazard functions. One may be forced to fit

$$\lambda_y(t|Z_i, W_i) = \lambda_0^{Y*}(t) + \beta_y^{*T} X_i, \tag{4.59}$$

where $\lambda_0^{Y*}(t)$ is an arbitrary baseline hazard function. Under the model (4.57),

$$\begin{aligned}\int_{-\infty}^{\infty} f(t|Z_i, W_i, u)dF(u|Z_i, W_i) = \{\lambda_0^Y(t) + \beta_y^T X_i\} \int_{-\infty}^{\infty} S(t|Z_i, W_i, u)dF(u|Z_i, W_i) \\ +\beta_{yu} \int_{-\infty}^{\infty} u \exp\{-\Lambda_0^Y(t) - t(\beta_y^T X_i + \beta_{yu} U_i)\} dF(u|Z_i, W_i),\end{aligned} \tag{4.60}$$

$$\int_{-\infty}^{\infty} S(t|Z_i, W_i, u)dF(u|Z_i, W_i) = \int_{-\infty}^{\infty} \exp\{-\Lambda_0^Y(t) - t(\beta_y^T X_i + \beta_{yu} U_i)\} dF(u|Z_i, W_i) \tag{4.61}$$

Thus,

$$\lambda_y(t|Z_i, W_i) = \lambda_0^Y(t) + \beta_y^T X_i + \beta_{yu} a(t|Z_i, W_i), \tag{4.62}$$

where

$$a(t|Z_i, W_i) = \frac{\int_{-\infty}^{\infty} u \exp\{-\Lambda_0^Y(t) - t(\beta_y^T X_i + \beta_{yu} U_i)\} dF(u|Z_i, W_i)}{\int_{-\infty}^{\infty} \exp\{-\Lambda_0^Y(t) - t(\beta_y^T X_i + \beta_{yu} U_i)\} dF(u|Z_i, W_i)}. \tag{4.63}$$

We can show that $a(t|Z_i, W_i) = \mu(Z_i, W_i) - t\beta_{yu}$. It then follows that

$$\lambda_y(t|Z_i, W_i) = \lambda_0^Y(t) - t\beta_{yu}^2 + \beta_{yu}\mu(0, W_i) + \{\beta_{yz} + \beta_{yu}\delta(Z_i, W_i)\} Z_i + \beta_{yw} W_i. \tag{4.64}$$

This implies that

$$\beta_{yz} = \beta_{yz}^* - \beta_{yu}\delta(Z_i, W_i), \tag{4.65}$$

which is the same as (4.16). Employing the negative control outcome with the additive model assumption of $\delta(Z_i, W_i)$ gives the previous formula (4.20).

4.2.3 Conditional likelihood approach

The sensitivity analysis of Lin et al. (1998) in the logistic regression for binary outcome and the proportional hazard model for survival time outcome appeared to be reliable if event rates are rare or $|\beta_{yu}|$ is small. In their simulation studies, in case of normal U_i, $\hat{\beta}_{yz}(\beta_{yu}, \delta)$ for logistic regression appeared to be accurate if the overall event rate is below 10% or $|\beta_{yu}| < 0.75$. The smaller δ gave less bias of $\hat{\beta}_{yz}(\beta_{yu}, \delta)$. For proportional hazard model, $\hat{\beta}_{yz}(\beta_{yu}, \delta)$ appeared to be reliable if the censoring percentage is greater than 90% or $|\beta_{yu}| < 0.75$. However, in many cases, those conditions may not be satisfied.

We proposed conditional likelihood approach for logistic regression model for binary outcome and proportional hazard model for survival outcome. Our method uses a negative control outcome to make the condition that the regression parameter for the unmeasured confounder should be small more likely to be met. The proposed method is applicable when $|\beta_{yu} - \beta_{nu}|$ is much less than $|\beta_{yu}|$, which will happen if β_{yu} and β_{nu} have the same signs and $|\beta_{nu}| \approx |\beta_{yu}|$. The key step is that first we construct conditional likelihood with a negative control outcome and apply the method of Lin et al. (1998) to the conditional likelihood.

As before, for binary outcomes, we use the logistic regression models,

$$\text{logit}\{P(Y_i = 1|Z_i, W_i, U_i)\} = \beta_y^T X_i + \beta_{yu} U_i, \tag{4.66}$$

$$\text{logit}\{P(N_i = 1|Z_i, W_i, U_i)\} = \beta_n^T X_i + \beta_{nu} U_i. \tag{4.67}$$

Let $S_i = Y_i + N_i$ be the sum of the two outcomes and $S = (S_1, ..., S_n)^T$. The distribution of (Y_i, N_i) conditional on $S_i = 1$ is given by

$$P(Y_i, N_i|S_i = 1) = \begin{cases} \frac{\exp(\theta^T X_i + \delta_u U_i)}{1+\exp(\theta^T X_i + \delta_u U_i)}, & \text{if } Y_i = 1, N_i = 0, \\ \frac{1}{1+\exp(\theta^T X_i + \delta_u U_i)}, & \text{if } Y_i = 0, N_i = 1, \end{cases} \tag{4.68}$$

where $\theta^T = (\delta_o, \beta_{yz}, \delta_w^T)$, $\delta_o = \beta_{yo} - \beta_{no}$, $\delta_w = \beta_{yw} - \beta_{nw}$ and $\delta_u = \beta_{yu} - \beta_{nu}$. We assume that $|\delta_u|$ is small, which is more acceptable than that $|\beta_{yu}|$ is small. The corresponding joint probability function to (4.68) is given by

$$L(\theta, \delta_u) = \prod_{i \in A} \left(\frac{\exp(\theta^T X_i + \delta_u U_i)}{1 + \exp(\theta^T X_i + \delta_u U_i)} \right)^{Y_i} \left(\frac{1}{1 + \exp(\theta^T X_i + \delta_u U_i)} \right)^{1-Y_i}, \tag{4.69}$$

where $A = \{i : S_i = 1\}$. We can see that (4.69) is the one for the following conditional logistic regression model,

$$\text{logit}\{\text{pr}(Y_i = 1 | Z_i, W_i, U_i, S_i = 1)\} = \theta^T X_i + \delta_u U_i. \tag{4.70}$$

One is forced to fit the conditional logistic regression model with measured variables (Z_i, W_i):

$$\text{logit}\{\text{pr}(Y_i = 1 | Z_i, W_i, S_i = 1)\} = \theta^{*T} X_i, \tag{4.71}$$

where $\theta^{*T} = (\delta_o^*, \beta_{yz}^*, \delta_w^{*T})$. Under the model (4.70),

$$\text{pr}(Y_i = 1 | Z_i, W_i, S_i = 1) = \exp(\theta^T X_i) \int_{-\infty}^{\infty} \frac{\exp(\delta_u u)}{1 + \exp(\theta^T X_i + \delta_u u)} dF(u | Z_i, W_i, S_i = 1) \tag{4.72}$$

In section 2.2, we assumed that the conditional distribution of U_i given (Z_i, W_i) is normal with mean $\eta(Z_i, W_i)$ and unit variance. Since $|\delta_u|$ is small, the same conditional distribution may approximately hold for U_i given $(Z_i, W_i, S_i = 1)$ because Y_i (or N_i) is approximately independent of U_i conditional on $(Z_i, W_i, S_i = 1)$ when $|\delta_u| \approx 0$, and hence S_i and U_i are also approximately independent. Straightforward algebraic calculation shows that

$$\text{logit}\{\text{pr}(Y_i = 1 | Z_i, W_i, S_i = 1)\} = \theta^T X_i + q(Z_i, W_i), \tag{4.73}$$

where

$$
\begin{aligned}
q(Z_i, W_i) =& \delta_u \eta(Z_i, W_i) + 0.5\delta_u^2 \\
& + \log \frac{\int_{-\infty}^{\infty} \{1 + \exp(\delta_u^2 + \theta^T X_i + \delta_u u)\}^{-1} \exp\{-\frac{u - \eta(Z_i, W_i)^2}{2}\} du}{\int_{-\infty}^{\infty} \{1 + \exp(\theta^T X_i + \delta_u u)\}^{-1} \exp\{-\frac{u - \eta(Z_i, W_i)^2}{2}\} du}.
\end{aligned} \tag{4.74}
$$

Since δ_u is small, $q(Z_i, W_i) \approx \delta_u \eta(Z_i, W_i) + 0.5\delta_u^2$. It then follows that

$$\beta_{yz} = \beta_{yz}^* - \delta_u \{\eta(1, W_i) - \eta(0, W_i)\} \tag{4.75}$$

Assume that the effects of Z_i and W_i are additive on $q(Z_i, W_i)$, i.e., $\eta(Z_i, W_i) = \eta_z(Z_i) + \eta_w(W_i)$. Then (4.75) reduces to

$$\beta_{yz} = \beta_{yz}^* - \delta_u \nu, \tag{4.76}$$

where $\nu = \eta_z(1) - \eta_z(0)$. The estimator of β_{yz}^* is obtained by fitting the conditional logistic regression model (4.71).

Now we describe the conditional likelihood method for survival data. Let T_{yi} and T_{ni} denote the main and negative control survival outcomes. We assume the following proportional hazard models

$$\lambda_y(t|Z_i, W_i, U_i) = \lambda_{0i}(t) \exp(\beta_y^T X_i + \beta_{yu} U_i) \tag{4.77}$$

$$\lambda_n(t|Z_i, W_i, U_i) = \lambda_{0i}(t) \exp(\beta_n^T X_i + \beta_{nu} U_i), \tag{4.78}$$

where different subject can have a different arbitrary baseline hazard function $\lambda_{0i}(t)$, but the pair within a subject shares the same baseline hazard function. First we suppose that there is no censoring. Under models (4.77) and (4.78), pair ranks are marginally sufficient for (θ, δ_u) (Holt and Prentice, 1974). The marginal likelihood for the parameters is proportional to the product of the following terms over pair

$$\mathrm{pr}(T_{yi} < T_{ni}) = \frac{\exp(\theta^T X_i + \delta_u U_i)}{1 + \exp(\theta^T X_i + \delta_u U_i)}, \tag{4.79}$$

$$\mathrm{pr}(T_{yi} > T_{ni}) = \frac{1}{1 + \exp(\theta^T X_i + \delta_u U_i)}. \tag{4.80}$$

Let Y_i denote the event $I(T_{yi} < T_{ni})$. The marginal likelihood can be written as

$$L_1(\theta, \delta_u | Z_i, W_i, U_i) \propto \prod_{i=1}^{n} \left\{ \frac{\exp(\theta^T X_i + \delta_u U_i)}{1 + \exp(\theta^T X_i + \delta_u U_i)} \right\}^{Y_i} \left\{ \frac{1}{1 + \exp(\theta^T X_i + \delta_u U_i)} \right\}^{1-Y_i}, \tag{4.81}$$

which corresponds to the logistic regression model,

$$\mathrm{logit}\{\mathrm{pr}(Y_i = 1 | Z_i, W_i, U_i)\} = \theta^T X_i + \delta_u U_i \tag{4.82}$$

Thus sensitivity analysis can be performed by the method of Lin et al. (1998) for the logistic regression.

Holt and Prentice (1974) pointed out that in the presence of censoring, marginally sufficient statistics generally depends on the censoring times and nuisance parameter. An important special case arises when members of a pair share the same censoring time and in this case the same form of marginal likelihood as (4.81) can be constructed. This may frequently occur in our setting because the two outcomes come from the same subject. A pair rank is observed if paired survivals are uncensored or singly censored but is unknown for doubly censored pairs.

Suppose that T_i^0 is the censoring time for the ith subject. Then

$$\mathrm{pr}(T_{yi} < T_{ni} | T_{yi} < T_i^0 \text{ or } T_{ni} < T_i^0) = \frac{\exp(\theta^T X_i + \delta_u U_i)}{1 + \exp(\theta^T X_i + \delta_u U_i)}, \tag{4.83}$$

$$\mathrm{pr}(T_{yi} > T_{ni} | T_{yi} < T_i^0 \text{ or } T_{ni} < T_i^0) = \frac{1}{1 + \exp(\theta^T X_i + \delta_u U_i)}, \tag{4.84}$$

which are equivalent to those for the uncensored case. The inference for the parameter can be made from

$$L_2(\theta, \delta_u | Z_i, W_i, U_i) \propto \prod_{i=1}^{r} \left\{ \frac{\exp(\theta^T X_i + \delta_u U_i)}{1 + \exp(\theta^T X_i + \delta_u U_i)} \right\}^{Y_i} \left\{ \frac{1}{1 + \exp(\theta^T X_i + \delta_u U_i)} \right\}^{1-Y_i}, \quad (4.85)$$

where pairs $i = 1, ..., r$ are uncensored or singly censored while pairs $i = r + 1, ..., n$ are doubly censored.

For general censoring scheme, the inference for (θ, δ_u) can be made with more restrictive pair ranks. That is, a pair rank is observed if both paired survivals are uncensored or singly censored but the censored part is greater than the uncensored part. Let T_{yi}^0 and T_{ni}^0 denote the censoring times for T_{yi} and T_{ni} respectively. Then

$$\text{pr}(T_{yi} < T_{ni} | i \in C) = \frac{\exp(\theta^T X_i + \delta_u U_i)}{1 + \exp(\theta^T X_i + \delta_u U_i)}, \quad (4.86)$$

$$\text{pr}(T_{yi} > T_{ni} | i \in C) = \frac{1}{1 + \exp(\theta^T X_i + \delta_u U_i)}, \quad (4.87)$$

where $C = C_1 \cup C_2 \cup C_3$, and $C_1 = \{i : T_{yi} < T_{yi}^0 \text{ and } T_{ni} < T_{ni}^0\}$, $C_2 = \{i : T_{yi} < T_{yi}^0 \text{ and } T_{ni} \geq T_{ni}^0 \text{ and } T_{ni}^0 \geq T_{yi}\}$ and $C_3 = \{i : T_{ni} < T_{ni}^0 \text{ and } T_{yi} \geq T_{yi}^0 \text{ and } T_{yi}^0 \geq T_{ni}\}$.

The inference for the parameter can be made from

$$L_3(\theta, \delta_u | Z_i, W_i, U_i) \propto \prod_{i=1}^{k} \left\{ \frac{\exp(\theta^T X_i + \delta_u U_i)}{1 + \exp(\theta^T X_i + \delta_u U_i)} \right\}^{Y_i} \left\{ \frac{1}{1 + \exp(\theta^T X_i + \delta_u U_i)} \right\}^{1-Y_i}, \quad (4.88)$$

where pairs $i = 1, ..., k$ are uncensored or singly censored but the censored part is greater than the uncensored part while pairs $i = k + 1, ..., n$ are doubly censored or singly censored with the censored part being less than the uncensored part.

4.3 Simulations

The goal of our simulation study is to examine the proposed methods based on marginal model and conditional likelihood using a negative control outcome and compared those to Lin's methods. We used logistic regression model for binary outcome and Cox proportional hazard model for survival outcomes.

First, we generated the exposure Z from Bernoulli distribution with an event probability of 0.5. Next, we generated the confounder W from $N(Z, 1)$ and the unmeasured confounder U from $N(Z + W, 1)$. We used the following logistic regression models for binary outcomes:

$$\text{logit}\{P(Y = 1|Z, W, U)\} = \beta_{yz}Z + \beta_{yw}W + aU, \tag{4.89}$$

$$\text{logit}\{P(N = 1|Z, W, U)\} = \beta_{nz}Z + \beta_{nw}W + bU. \tag{4.90}$$

For survival outcomes, we used the following proportional hazard models:

$$\lambda_y(t|Z, W, U) = \exp(\beta_{yz}Z + \beta_{yw}W + aU), \tag{4.91}$$

$$\lambda_n(t|Z, W, U) = \exp(\beta_{nz}Z + \beta_{nw}W + bU). \tag{4.92}$$

The censoring variables for the main and the negative control survival outcomes were generated from Unif$(0, 4)$ separately, and thus for conditional likelihood approach we used the method with general censoring scheme. With the parameter values of (a, b) we used, this simulation configuration gives the censoring rates of about 20% and 21% for the main and the negative control survival outcomes respectively. For each simulation scenario of the binary and survival outcomes, we generated 500 data sets with the sample size of 1600.

The regression parameter vector $(\beta_{yz}, \beta_{nz}, \beta_{yw}, \beta_{nw})$ was set to be $(1, 0, 1, 1)$. For (a, b), we used $(0.5, 0.4)$ and $(1, 0.8)$. If $(a, b) = (0.5, 0.4)$, then the condition ($|\beta_{yu}| < 0.75$), for which the approximations of Lin's methods used for binary and proportional hazard models work well, is satisfied. The case where $(a, b) = (1, 0.8)$ does not satisfy this condition, and thus the methods of Lin et al. (1998) may not work well in this setting. Those two vectors all give $a/b = 1.25$.

To implement the methods of Lin et al. (1998), we used the formula $\hat{\beta}_{yz}(\beta_{yu}, \delta)$ in (4.18) with varying β_{yu} and δ, where $\hat{\beta}_{yz}$ was obtained by fitting model (4.89) or (4.91) without U. To implement the marginal model methods, we used the formula $\hat{\beta}_{yz} - \gamma_u\hat{\beta}_{nz}$ with varying γ_u, where $\hat{\beta}_{nz}$ was obtained by fitting model (4.90) or (4.92) without U. Lastly, for the conditional

Table 4.1: Sensitivity parameter values used in simulations.

Parameter	(a, b)	Values
β_{yu}	$(0.5, 0.4)$	0.3, 0.4, 0.5, 0.6, 0.7
	$(1.0, 0.8)$	0.8, 0.9, 1.0, 1.1, 1.2
δ_u	$(0.5, 0.4)$	0.01, 0.05, 0.10, 0.15, 0.20
	$(1.0, 0.8)$	0.01, 0.10, 0.20, 0.30, 0.40
δ, ν		0.50, 0.75, 1.00, 1.25, 1.50
γ_u		0.50, 0.53, 0.57, 0.60, 0.64, 0.67, 0.71, 0.76, 0.80, 0.85, 0.89, 0.95, 1.00, 1.06, 1.12, 1.18, 1.25, 1.32, 1.40, 1.48, 1.57, 1.67, 1.77, 1.88, 2.00

likelihood methods, we applied the methods of Lin et al. (1998) to the conditional likelihoods with varying δ_u and ν, where δ_u and ν play roles as β_{yu} and δ. Since β_{yu} and δ_u need to be around the values of a and $a - b$ for reliable sensitivity analysis, those parameter values were varied depending on the values of a and $a - b$ respectively. To be specific, we varied β_{yu} within $\{0.3, 0.7\}$ when $a = 0.5$ and within $\{0.8, 1.2\}$ when $a = 1.0$. We varied δ_u within $\{0.01, 0.2\}$ when $a - b = 0.1$ and within $\{0.01, 0.40\}$ when $a - b = 0.2$. All of the parameter values of $(\beta_{yu}, \delta_u, \delta, \nu, \gamma_u)$ considered were listed in Table 4.1.

Table 4.2-4.5 show the simulation results. For each method, we calculated the bias, and the empirical coverage rate (ECR) of a 95% confidence interval. In each table, the bold characters indicate the results when the sensitivity parameters have the same value as those of the corresponding parameters. It may be reasonable to say that the method having a less bias and an ECR close to 95% at or near the true parameter values works well.

In the logistic regression with $(a, b) = (0.5, 0.4)$, all of the three methods worked well. The ECR of Lin's method was slightly lower than 95%, however, conditional likelihood and marginal model methods gave correct coverage probabilities. In the logistic regression with $(a, b) = (1.0, 0.8)$, the only conditional likelihood method gave correct inference results at the true value. Lin's method failed to give correct inference results at the true value because of huge unmeasured confounding. Since the proposed marginal model method uses the same approximations used in Lin's method, it also failed to yield correct results like Lin's, however,

Table 4.2: Logistic regression model with (a, b) = $(0.5, 0.4)$

Lin's method				Marginal model				Conditional Likelihood			
(β_{yu},δ)	Bias	ECR	LEN	γ_u	Bias	ECR	LEN	(δ_u, ν)	Bias	ECR	LEN
(0.30,0.50)	0.294	0.530	0.608	0.50	0.252	0.668	0.654	(0.01,0.50)	0.096	0.936	0.856
(0.30,0.75)	0.219	0.702		0.53	0.240	0.702	0.660	(0.01,0.75)	0.094	0.936	
(0.30,1.00)	0.144	0.848		0.57	0.227	0.728	0.667	(0.01,1.00)	0.091	0.938	
(0.30,1.25)	0.069	0.930		0.60	0.214	0.760	0.675	(0.01,1.25)	0.089	0.938	
(0.30,1.50)	-0.006	0.930		0.64	0.200	0.784	0.684	(0.01,1.50)	0.086	0.940	
(0.40,0.50)	0.244	0.648		0.67	0.185	0.808	0.693	(0.05,0.50)	0.076	0.938	
(0.40,0.75)	0.144	0.848		0.71	0.170	0.838	0.703	(0.05,0.75)	0.064	0.936	
(0.40,1.00)	0.044	0.948		0.76	0.154	0.860	0.715	(0.05,1.00)	0.051	0.936	
(0.40,1.25)	-0.056	0.910		0.80	0.137	0.876	0.727	(0.05,1.25)	0.039	0.936	
(0.40,1.50)	-0.156	0.798		0.85	0.120	0.904	0.741	(0.05,1.50)	0.026	0.938	
(0.50,0.50)	0.194	0.760		0.89	0.101	0.910	0.756	(0.10,0.50)	0.051	0.936	
(0.50,0.75)	0.069	0.930		0.95	0.081	0.922	0.772	(0.10,0.75)	0.026	0.938	
(0.50,1.00)	**-0.056**	**0.910**		1.00	0.061	0.928	0.790	**(0.10,1.00)**	**0.001**	**0.934**	
(0.50,1.25)	-0.181	0.758		1.06	0.039	0.932	0.809	(0.10,1.25)	-0.024	0.934	
(0.50,1.50)	-0.306	0.498		1.12	0.015	0.934	0.831	(0.10,1.50)	-0.049	0.926	
(0.60,0.50)	0.144	0.848		1.18	-0.009	0.936	0.854	(0.15,0.50)	0.026	0.938	
(0.60,0.75)	-0.006	0.930		**1.25**	**-0.035**	**0.936**	**0.879**	(0.15,0.75)	-0.011	0.936	
(0.60,1.00)	-0.156	0.798		1.32	-0.063	0.930	0.907	(0.15,1.00)	-0.049	0.926	
(0.60,1.25)	-0.306	0.498		1.40	-0.093	0.918	0.938	(0.15,1.25)	-0.086	0.916	
(0.60,1.50)	-0.456	0.182		1.48	-0.124	0.908	0.971	(0.15,1.50)	-0.124	0.906	
(0.70,0.50)	0.094	0.912		1.57	-0.158	0.892	1.008	(0.20,0.50)	0.001	0.934	
(0.70,0.75)	-0.081	0.876		1.67	-0.195	0.876	1.048	(0.20,0.75)	-0.049	0.926	
(0.70,1.00)	-0.256	0.610		1.77	-0.234	0.856	1.093	(0.20,1.00)	-0.099	0.912	
(0.70,1.25)	-0.431	0.226		1.88	-0.277	0.834	1.142	(0.20,1.25)	-0.149	0.880	
(0.70,1.50)	-0.606	0.030		2.00	-0.323	0.810	1.195	(0.20,1.50)	-0.199	0.840	

Table 4.3: Logistic regression model with (a, b) = $(1.0, 0.8)$

Lin's method				Marginal model				Conditional Likelihood			
(β_{yu},δ)	Bias	ECR	LEN	γ_u	Bias	ECR	LEN	(δ_u, ν)	Bias	ECR	LEN
(0.80,0.50)	0.343	0.482	0.673	0.50	0.384	0.424	0.702	(0.01,0.50)	0.150	0.908	0.986
(0.80,0.75)	0.143	0.878		0.53	0.361	0.476	0.707	(0.01,0.75)	0.147	0.910	
(0.80,1.00)	-0.057	0.916		0.57	0.337	0.524	0.714	(0.01,1.00)	0.145	0.910	
(0.80,1.25)	-0.257	0.668		0.60	0.313	0.590	0.721	(0.01,1.25)	0.142	0.912	
(0.80,1.50)	-0.457	0.256		0.64	0.286	0.658	0.729	(0.01,1.50)	0.140	0.914	
(0.90,0.50)	0.293	0.618		0.67	0.259	0.716	0.737	(0.10,0.50)	0.105	0.934	
(0.90,0.75)	0.068	0.936		0.71	0.231	0.772	0.747	(0.10,0.75)	0.080	0.936	
(0.90,1.00)	-0.157	0.840		0.76	0.201	0.828	0.758	(0.10,1.00)	0.055	0.934	
(0.90,1.25)	-0.382	0.394		0.80	0.169	0.858	0.770	(0.10,1.25)	0.030	0.930	
(0.90,1.50)	-0.607	0.068		0.85	0.136	0.884	0.783	(0.10,1.50)	0.005	0.942	
(1.00,0.50)	0.243	0.724		0.89	0.101	0.914	0.797	(0.20,0.50)	0.055	0.934	
(1.00,0.75)	-0.007	0.922		0.95	0.065	0.936	0.813	(0.20,0.75)	0.005	0.942	
(1.00,1.00)	**-0.257**	**0.668**		1.00	0.026	0.946	0.831	**(0.20,1.00)**	**-0.045**	**0.938**	
(1.00,1.25)	-0.507	0.192		1.06	-0.015	0.940	0.850	(0.20,1.25)	-0.095	0.918	
(1.00,1.50)	-0.757	0.022		1.12	-0.059	0.916	0.872	(0.20,1.5)	-0.145	0.890	
(1.10,0.50)	0.193	0.802		1.18	-0.105	0.910	0.895	(0.30,0.50)	0.005	0.942	
(1.10,0.75)	-0.082	0.906		**1.25**	**-0.153**	**0.890**	**0.921**	(0.30,0.75)	-0.070	0.922	
(1.10,1.00)	-0.357	0.442		1.32	-0.205	0.856	0.950	(0.30,1.00)	-0.145	0.890	
(1.10,1.25)	-0.632	0.054		1.40	-0.261	0.820	0.981	(0.30,1.25)	-0.220	0.828	
(1.10,1.50)	-0.907	0.002		1.48	-0.320	0.774	1.016	(0.30,1.50)	-0.295	0.764	
(1.20,0.50)	0.143	0.878		1.57	-0.384	0.714	1.054	(0.40,0.50)	-0.045	0.938	
(1.20,0.75)	-0.157	0.840		1.67	-0.452	0.612	1.096	(0.40,0.75)	-0.145	0.890	
(1.20,1.00)	-0.457	0.256		1.77	-0.526	0.554	1.142	(0.40,1.00)	-0.245	0.806	
(1.20,1.25)	-0.757	0.022		1.88	-0.605	0.480	1.193	(0.40,1.25)	-0.345	0.702	
(1.20,1.50)	-1.057	0.000		2.00	-0.691	0.432	1.250	(0.40,1.50)	-0.445	0.572	

Table 4.4: Proportional hazard model with (a, b) = $(0.5, 0.4)$

Lin's method				Marginal model				Conditional Likelihood			
(β_{yu},δ)	Bias	ECR	LEN	γ_u	Bias	ECR	LEN	(δ_u, ν)	Bias	ECR	LEN
(0.30,0.50)	0.169	0.258	0.258	0.50	0.139	0.496	0.285	(0.01,0.50)	0.100	0.882	0.508
(0.30,0.75)	0.094	0.702		0.53	0.128	0.576	0.288	(0.01,0.75)	0.098	0.884	
(0.30,1.00)	0.019	0.916		0.57	0.116	0.662	0.292	(0.01,1.00)	0.095	0.888	
(0.30,1.25)	-0.056	0.832		0.60	0.103	0.730	0.296	(0.01,1.25)	0.093	0.894	
(0.30,1.50)	-0.131	0.494		0.64	0.090	0.788	0.301	(0.01,1.50)	0.090	0.894	
(0.40,0.50)	0.119	0.558		0.67	0.077	0.832	0.305	(0.05,0.50)	0.080	0.912	
(0.40,0.75)	0.019	0.916		0.71	0.062	0.888	0.311	(0.05,0.75)	0.068	0.920	
(0.40,1.00)	-0.081	0.756		0.76	0.047	0.932	0.316	(0.05,1.00)	0.055	0.936	
(0.40,1.25)	-0.181	0.232		0.80	0.032	0.954	0.323	(0.05,1.25)	0.043	0.946	
(0.40,1.50)	-0.281	0.016		0.85	0.015	0.964	0.330	(0.05,1.50)	0.030	0.950	
(0.50,0.50)	0.069	0.804		0.89	-0.003	0.962	0.337	(0.10,0.50)	0.055	0.936	
(0.50,0.75)	-0.056	0.832		0.95	-0.021	0.958	0.345	(0.10,0.75)	0.030	0.950	
(0.50,1.00)	**-0.181**	**0.232**		1.00	-0.040	0.930	0.354	**(0.10,1.00)**	**0.005**	**0.960**	
(0.50,1.25)	-0.306	0.010		1.06	-0.061	0.904	0.364	(0.10,1.25)	-0.020	0.964	
(0.50,1.50)	-0.431	0.000		1.12	-0.083	0.858	0.374	(0.10,1.50)	-0.045	0.938	
(0.60,0.50)	0.019	0.916		1.18	-0.106	0.818	0.386	(0.15,0.50)	0.030	0.950	
(0.60,0.75)	-0.131	0.494		**1.25**	**-0.130**	**0.764**	**0.398**	(0.15,0.75)	-0.007	0.960	
(0.60,1.00)	-0.281	0.016		1.32	-0.156	0.680	0.412	(0.15,1.00)	-0.045	0.938	
(0.60,1.25)	-0.431	0.000		1.40	-0.184	0.612	0.427	(0.15,1.25)	-0.082	0.908	
(0.60,1.50)	-0.581	0.000		1.48	-0.214	0.526	0.443	(0.15,1.50)	-0.120	0.846	
(0.70,0.50)	-0.031	0.884		1.57	-0.246	0.440	0.461	(0.20,0.50)	0.005	0.960	
(0.70,0.75)	-0.206	0.124		1.67	-0.280	0.390	0.480	(0.20,0.75)	-0.045	0.938	
(0.70,1.00)	-0.381	0.000		1.77	-0.317	0.306	0.502	(0.20,1.00)	-0.095	0.886	
(0.70,1.25)	-0.556	0.000		1.88	-0.357	0.252	0.525	(0.20,1.25)	-0.145	0.784	
(0.70,1.50)	-0.731	0.000		2.00	-0.400	0.210	0.550	(0.20,1.50)	-0.195	0.678	

Table 4.5: Proportional hazard model with (a, b) = $(1.0, 0.8)$

Lin's method				Marginal model				Conditional Likelihood			
$(\beta_{yu}$,$\delta)$	Bias	ECR	LEN	γ_u	Bias	ECR	LEN	(δ_u, ν)	Bias	ECR	LEN
(0.80,0.50)	-0.023	0.892	0.258	0.50	0.076	0.826	0.285	(0.01,0.50)	0.181	0.722	0.520
(0.80,0.75)	-0.223	0.112		0.53	0.057	0.892	0.289	(0.01,0.75)	0.178	0.730	
(0.80,1.00)	-0.423	0.000		0.57	0.037	0.932	0.293	(0.01,1.00)	0.176	0.732	
(0.80,1.25)	-0.623	0.000		0.60	0.016	0.958	0.297	(0.01,1.25)	0.173	0.736	
(0.80,1.50)	-0.823	0.000		0.64	-0.006	0.968	0.301	(0.01,1.50)	0.171	0.744	
(0.90,0.50)	-0.073	0.758		0.67	-0.029	0.958	0.306	(0.10,0.50)	0.136	0.836	
(0.90,0.75)	-0.298	0.014		0.71	-0.053	0.916	0.311	(0.10,0.75)	0.111	0.892	
(0.90,1.00)	-0.523	0.000		0.76	-0.078	0.836	0.317	(0.10,1.00)	0.086	0.920	
(0.90,1.25)	-0.748	0.000		0.80	-0.104	0.762	0.324	(0.10,1.25)	0.061	0.938	
(0.90,1.50)	-0.973	0.000		0.85	-0.132	0.660	0.331	(0.10,1.50)	0.036	0.946	
(1.00,0.50)	-0.123	0.528		0.89	-0.161	0.542	0.338	(0.20,0.50)	0.086	0.920	
(1.00,0.75)	-0.373	0.000		0.95	-0.192	0.406	0.346	(0.20,0.75)	0.036	0.946	
(1.00,1.00)	**-0.623**	**0.000**		1.00	-0.225	0.312	0.355	**(0.20,1.00)**	**-0.014**	**0.954**	
(1.00,1.25)	-0.873	0.000		1.06	-0.259	0.204	0.365	(0.20,1.25)	-0.064	0.938	
(1.00,1.50)	-1.123	0.000		1.12	-0.295	0.096	0.376	(0.20,1.50)	-0.114	0.852	
(1.10,0.50)	-0.173	0.296		1.18	-0.334	0.050	0.387	(0.30,0.50)	0.036	0.946	
(1.10,0.75)	-0.448	0.000		**1.25**	**-0.375**	**0.020**	**0.400**	(0.30,0.75)	-0.039	0.954	
(1.10,1.00)	-0.723	0.000		1.32	-0.418	0.008	0.414	(0.30,1.00)	-0.114	0.852	
(1.10,1.25)	-0.998	0.000		1.40	-0.465	0.004	0.429	(0.30,1.25)	-0.189	0.684	
(1.10,1.50)	-1.273	0.000		1.48	-0.515	0.000	0.445	(0.30,1.50)	-0.264	0.502	
(1.20,0.50)	-0.223	0.112		1.57	-0.568	0.000	0.463	(0.40,0.50)	-0.014	0.954	
(1.20,0.75)	-0.523	0.000		1.67	-0.625	0.000	0.482	(0.40,0.75)	-0.114	0.852	
(1.20,1.00)	-0.823	0.000		1.77	-0.687	0.000	0.504	(0.40,1.00)	-0.214	0.638	
(1.20,1.25)	-1.123	0.000		1.88	-0.753	0.000	0.527	(0.40,1.25)	-0.314	0.360	
(1.20,1.50)	-1.423	0.000		2.00	-0.826	0.000	0.553	(0.40,1.50)	-0.414	0.106	

the results were much better than those of Lin's with less bias and more accurate ECR at the true value. In the proportional hazard model, the only conditional likelihood method worked well for both $(a, b) = (0.5, 0.4)$ and $(a, b) = (1.0, 0.8)$.

The naive estimator for the exposure effect is $\hat{\beta}_{yz}$, which was obtained by fitting the outcome model ignoring U. The biases (ECRs) of $\hat{\beta}_{yz}$ were 0.444 (0.194), 0.743 (0.006), 0.319 (0.006)and 0.377 (0.000) for Table 4.2-4.5. The variance of Lin's estimator is equal to that of the naive estimator because it is obtained by adding a constant to the naive estimator.

The simulation results showed proposed methods had larger variances than Lin's methods. The marginal model method has an additional variability coming from $\hat{\beta}_{nu}$ and its variance is approximately proportional to γ_u^2. The conditional likelihood method uses fewer samples than the other methods because it uses the subjects that satisfies a certain condition.

CHAPTER 5: PRACTICABLE CONFIDENCE INTERVALS FOR CURRENT STATUS DATA

5.1 Introduction

Nonparametric estimation of binary isotonic regression functions has been well studied theoretically and its connection to current status survival data has been described. A systematic overview of available binary isotonic regression methods is given in Ghosh et al. (2008) Although non-parametric estimation of parameters can be easily obtained using the pooled-adjacent violators algorithm, confidence intervals and other assessments of sampling variability are rarely employed, limiting the practical utility of this methodology. Inferential difficulties arise because the nonparametric estimator converges at a rate that is slower than the usual parametric rate of $n^{1/2}$.

Bootstrapping is a popular technique for computing confidence intervals in settings where analytic formulae may not exist. However, the usual nonparametric bootstrap is not theoretically justified for binary isotonic regression. Ghosh et al. (2008) investigated three alternative inferential procedures for the nonparametric maximum likelihood estimator (NPMLE): (i) a Wald-based method; (ii) a subsampling-based method; and (iii) a likelihood-ratio test-based method. Their simulation study indicated that the LRT and sub-sampling methods have superior performance to the Wald method, most notably with small and moderate sample sizes, where the Wald-based intervals were found to perform poorly. Unfortunately, to our knowledge, there is no publicly available software for computing confidence intervals using either the LRT or the subsampling approach, owing in part to their computational complexities. This has limited their adoption in statistical practice.

In contrast, the Wald-type confidence intervals are straightforward to compute, involving only "smooth" estimation of certain density functions. We hypothesized that transformations of the estimated regression function would substantially improve the empirical coverage the Wald CIs. This approach is motivated by related work in survival analysis with right censored data. Bie et al. (1987) and Borgan and Liestol (1990) studied confidence intervals for the cumulative hazard function and the survival function, demonstrating numerically that CIs based on transformations of the Nelson-Aalen and Kaplan-Meier estimators may have better properties than untransformed CIs. In isotonic regression, the results are more complicated, owing to the facts that the estimators being transformed converge at a rate of $n^{1/3}$ and have non-normal distributions, which differs from the earlier survival settings.

In the present paper, we establish the distribution of the transformed NPMLE for binary isotonic regression and use this result to construct Wald CIs which are as easy to implement as the untransformed Wald CIs. We evaluate via extensive simulation studies the performance of these across a variety of data generating models. We also consider a simple nonparametric bootstrap procedure. Finally, we apply all of the inferential methods to two example studies: 1) a large study of age at menopause using cross-sectional survey data; and 2) a small study of time until tumor development in mice.

5.2 Nonparametric confidence interval methods

Let (S, T) be a pair of random variables where S is observable and T is not observable and distributed as G, and δ be an indicator for whether T is less than S or not ($\delta = 1$ if $T \leq S$, $\delta = 0$ otherwise). We assume the following nonparametric regression model:

$$\Pr(\delta|S) = G(S), \tag{5.1}$$

where G is assumed to be monotonically increasing and continuously differentiable on $[0, \infty)$ with $G(0) = 0$ and $\lim_{z\to\infty} G(z) = 1$. The model (1) is binary isotonic regression, where the effect of S on δ is assumed monotone but otherwise unspecified. In Ghosh et al. (2008), S is tumour size and δ is an indicator of tumour metastasis. In current status data setting, S is an observation time and δ indicates whether the failure time, T, is less than the observation time, S.

The data consist of n independent and identically distributed observations (δ_1, S_1), (δ_2, S_2), ..., (δ_n, S_n). The joint density of (δ, S) is $\{G(s)\}^{\delta}\{1-G(s)\}^{1-\delta}h(s)$, where $h(\cdot)$ is the density of function of S. Ignoring h, the likelihood function of the observed data is

$$L_n(\{\delta_i, s_i\}_{i=1}^n) = \prod_{i=1}^n \{G(s_i)\}^{\delta_i}\{1-G(s_i)\}^{1-\delta_i}. \tag{5.2}$$

Although we evaluate G at observed values of S, the standard estimation method for current status survival data (the NPMLE method) can estimate the distribution of T unconditionally. The estimate of G maximizing this likelihood function may be obtained using standard algorithms. Let $S_{(i)}$ be the ith smallest value of tumor size (or observation time) and $\delta_{(i)}$ be the corresponding indicator function for tumor metastasis (or the indicator for whether failure time is less than observation time). The NPMLE of G is the left derivative of the greatest convex minorant of the points $\{i, \sum_{j=0}^{i} \delta_{(i)}\}$. The theoretical properties of the NPMLE procedure are given in Groeneboom and Wellner (1992), including its consistency and limiting distribution. Interestingly, the variance of the NPMLE converges more slowly than the usual $n^{1/2}$ rate for parametric estimators and for nonparametric estimators in "regular" problems, eg, the Kaplan-Meier estimator. Moreover, appropriately standardized, the NPMLE does not have a standard normal distribution. This greatly complicates inference.

The asymptotic distribution of the NPMLE of $G(s_0), \hat{G}_n(s_0)$, is

$$n^{1/3}\{\hat{G}_n(s_0) - G(s_0)\} \to_d \left[\frac{4g(s_0)G(s_0)\{1-G(s_0)\}}{h(s_0)}\right]^{1/3} Z \equiv CZ, \tag{5.3}$$

where $g(s_0)$ is the derivative of G at fixed point s_0 and Z is the location of the minimum of $W(t) + t^2$, with W being the standard two-sided Brownian motion starting from 0. Note that the distribution of W does not depend on any unknown parameters and hence its percentiles may be easily tabulated. Hence, the main difficulty in constructing confidence intervals is the estimation of the scaling constant C in (3).

An asymptotic 95% Wald-based confidence interval for $G(s_0)$ using the limiting distribution of the NPMLE is:

$$\left\{\hat{G}_n(s_0) - n^{-1/3}\hat{C}_n \times .99818,\ \hat{G}_n(s_0) + n^{-1/3}\hat{C}_n \times .99818\right\}, \tag{5.4}$$

where the scaling factor C is estimated by

$$\hat{C}_n = \left[\frac{4\hat{g}_n(s_0)\hat{G}_n(s_0)\{1 - \hat{G}_n(s_0)\}}{\hat{h}_n(s_0)}\right]^{1/3}, \tag{5.5}$$

and $\hat{g}_n(s_0)$ and $\hat{h}_n(s_0)$ are the estimates of g and h at s_0. To obtain $\hat{C}_n$, we need to estimate $g(s_0)$ and $h(s_0)$. Since S is observable for all samples, nonparametric density estimation can be used to estimate $h(s_0)$. In this article, we use kernel density estimation for $h(s_0)$ with the bandwidth chosen by the data-based direct plug-in methodology for band width selection Sheather and Jones (1991). The estimation for $g(s_0)$ is more difficult than the estimation for $h(s_0)$, since the failure times are not known exactly with current status data. We used kernel smoothing of the NPMLE $\hat{G}_n$, where the bandwidth was chosen by likelihood-based cross-validation (Pan 2000, Banerjee and Wellner 2005). A standard normal kernel was used in estimating $g(s_0)$ and $h(s_0)$.

To improve the performance of the Wald methods, we now describe an approach based on constructing confidence intervals using the transformed NPMLE. Let $A(\cdot)$ be a strictly monotone function which is differentiable in a neighborhood of $G(s_0)$ and has inverse $A^{-1}(\cdot)$. A first order Taylor expansion, eg, the Delta method, gives that

$$n^{1/3}\left\{A\left(\hat{G}_n(s_0)\right) - A\left(G_n(s_0)\right)\right\} \to_d \dot{A}\left(G(s_0)\right)CZ, \tag{5.6}$$

where $\dot{A}$ is the derivative of A. The asymptotic 95% Wald-based confidence interval for $A\left(G_n(s_0)\right)$ is constructed as follows:

$$A\left(\hat{G}_n(s_0)\right) \pm n^{-1/3}\dot{A}\left(\hat{G}_n(s_0)\right)\hat{C}_n \times .99818. \tag{5.7}$$

As with right censored data (Bie et al. 1987, Borgan and Liestol 1990), one might expect that with current status data, such transformed intervals will perform better than the untransformed intervals.

It is worth noting that no additional quantities need to be estimated, beyond those needed for the untransformed Wald interval with $A(u) = u$. In this article, we consider two transformations, $A(x) = \log(x/(1-x))$ and $A(x) = \log(-\log(x))$, which are the logit and log(-log) transformations. These transformations have previously been studied when constructing transformed Wald confidence intervals with right censored survival data (Bie et al. 1987, Borgan and Liestol 1990). The derivatives of $A(\cdot)$ for the logit and log(-log) transformations are $\dot{A}(x) = 1/(x(1-x))$ and $\dot{A}(x) = 1/(x\log(x))$, respectively.

Let (L, U) denote the lower and upper confidence limits of $A(G(s_0))$. The logit transformation gives the following 95% confidence interval for $G(s_0)$:

$$\left(\frac{e^L}{1+e^L}, \frac{e^U}{1+e^U}\right). \tag{5.8}$$

The log(-log) transformation gives the following 95% confidence interval for $G(s_0)$:

$$\left(e^{-e^U}, e^{-e^L}\right). \tag{5.9}$$

For general monotone increasing A, the 95% CI is $(A^{-1}(L), A^{-1}(U))$.

Next, we consider the nonparametric bootstrap method. Bootstrap samples of size n are constructed by resampling from the n observations $(\delta_i, S_i)_{i=1}^n$. In each bootstrap sample, one calculates the NPMLE of $G(s_0)$, $\hat{G_n}^*(s_0)$, using the bootstrap sample. We iterate this procedure a large number, B, times, say 500. A 95% confidence interval for $G(s_0)$ using the bootstrap is given by the 2.5 and 97.5 percentiles of the distribution of $\hat{G}_n^*(s_0)$.

To compute the Wald CI, we need to estimate the scaling factor C, which involves "smooth" estimation of $g(s_0)$ and $h(s_0)$, and hence may be unstable with small to moderate sample sizes. As an alternative, one may use the bootstrap to estimate C using the following approach. The asymptotic results for the NPMLE yield that the limiting distribution for Z is symmetric, with $E(Z^k) = 0$ for k odd, and the second absolute moment of Z is 0.2636 (Groeneboom and Wellner 2001). Thus the variance of Z is 0.2636. By eq. (3), asymptotically $Var(n^{1/3}\hat{G}_n(s_0)) = C^2 Var(Z) = C^2(0.2636)$, and hence C can be written as $C = n^{1/3}\sqrt{Var(\hat{G}_n(s_0))}/\sqrt{0.2636}$. The bootstrap estimator for C, which will be denoted as $\hat{C}^*$, is computed by estimating $Var(\hat{G}_n(s_0))$ based on bootstrap samples. The 95% bootstrap-Wald CI is given by

$$\hat{G}_n(s_0) \pm n^{-1/3}\hat{C}^* \times .99818. \tag{5.10}$$

For the transformed bootstrap-Wald CI, we estimated
$C_A = \dot{A}(\hat{G}_n(s_0))C$ by $\hat{C_A}^* = n^{1/3}\sqrt{\hat{Var}^*(A(\hat{G}_n(s_0)))}/\sqrt{0.2636}$, where $\hat{Var}^*(A(\hat{G}_n(s_0)))$ is the estimate of $Var(A(\hat{G}_n(s_0)))$ based on bootstrap. The 95% transformed bootstrap-Wald CI is given by

$$A^{-1}\left\{A(\hat{G}_n(s_0)) \pm n^{-1/3}\hat{C_A}^* \times .99818\right\}. \tag{5.11}$$

In addition to the confidence interval methods, we introduce a simple method to examine the reliability of the confidence interval obtained under given data. The basic idea is as follows. First, one generates data to mimic the given real data many times. Then, in each simulated dataset, one constructs confidence intervals as described above and computes performance measures, like empirical coverage probabilities, treating the simulation model as the truth. The sensitivity of the performance measures to the underlying simulation model may be assessed. We applied this approach to the two real data examples in Section 4, with the details reported in that section.

5.3 Simulations

The goal of our simulations is to compare the performance of the standard Wald-type confidence intervals to Wald-type CI based on transformations as well as confidence intervals based on the nonparametric bootstrap. For simulations, we adopted the same data generating mechanism as used in Banerjee and Wellner (2005) who studied confidence intervals for current status data.

This will facilitate comparison with the LRT and subsampling confidence intervals, which were examined in Banerjee and Wellner (2005).

We begin by presenting the results with $h(s) = e^{-s}$ and $G(s) = 1 - e^{-s}$, where both the observation time and the failure time are unit exponential. We compared the CIs at three points s_0 such that $G(s_0)$=0.2, 0.5 and 0.8. The sample sizes we consider are n=25, 50, 75, 100, 200, 500, 800, 1000 and 3000. We repeatedly generated data 5000 times and calculated the bias, the estimated asymptotic standard errors of $\hat{G}(s_0)$, average length of confidence intervals (Len) and coverage rate (CV). The bias is obtained as the average of $(\hat{G}_n(s_0) - G(s_0))$, $\hat{sd}(\hat{G}(s_0))$ is obtained by $n^{-1/3}\hat{C}\sqrt{Var(Z)}$, Len is calculated by the average of U−L, and the CV is the empirical coverage probability of the confidence intervals.

Tables 5.1 presents the simulation results for (S ~exp(1), T ~exp(1)) at the three points s_0 respectively (Tables 5.2 ~ 5.9 in the Appendix show the results for the rest of the simulation settings). As expected, as sample size increases, bias and average length of confidence intervals decrease, and coverage rate increases. For sample size n=3000, all of the methods produce satisfactory results in the sense that their coverage rates are near 0.95.

In the simulation setting (S ~exp(1), T ~exp(1)), Banerjee and Wellner (2005) compared the three methods outlined in Section 1 at $G(s_0)$=0.5. We compared our results and theirs for the Wald CIs (non-transformed) at $G(s_0)$=0.5, and confirmed that the two results are very similar, except that we obtained somewhat higher coverage rates for large sample sizes ($n \geq 500$). Given these similarities, it seems reasonable to compare our results for the Wald CIs using transformations and bootstrap to the simulation results in Banerjee and Wellner (2005) for the LRT and subsampling methods. Considering the result at $G(s_0) = 0.5$, the logit and log(-log) transformed Wald CIs produce results that are competitive with the likelihood-ratio-based (LRT) CI in Banerjee and Wellner (2005) in terms of average length and coverage probabilities. We note that the nonparametric bootstrap CI also produces comparable results to those from subsampling CIs in Banerjee and Wellner (2005) in terms of average length and coverage rate.

The main discovery in our simulations is that the Wald intervals may be improved by transformations across a wide range of observation and failure time scenarios. In each of our data generation models, one observes that the transformations improved the coverage of the Wald CI by 5-10$\%$ with small to moderate sample sizes. Comparing the logit and log(-log) transformations, there is some evidence that the logit transformation provides greater improvement than the log(-log) transformation. In addition, the numerical results indicate that the nonparametric bootstrap may have better coverage than the untransformed Wald CI in small to moderate sample sizes, but may not achieve the nominal coverage rate with large sample sizes. The results for the bootstrap Wald CI are similar to those for the nonparametric bootstrap.

The simulations provide evidence that the performances of the methods for constructing CIs are affected by the point s_0 where the confidence interval is constructed. For overall range of samples sizes and the distributions for S and T, the performances at $G(s_0)$ = 0.5 (s_0 is a median) are better than those at the other points, $G(s_0)$ = 0.2 and $G(s_0)$ = 0.8. Relative to G(s_0)=0.5 (s_0 is the median), the other time points are in the tails of the distribution. For these time points, $h(s_0)$ has a relatively small value when (S ~ $exp(1)$, $G(s_0)$ = 0.8) and S ~ $gamma(3, 1/3)$, and with small sample sizes, instability in kernel estimation may inflate the estimated value of C and make the confidence interval unstable. The diminished performances at $G(s_0)$ = 0.2 and $G(s_0)$ = 0.8 may also be related to the boundary problem in nonparametric estimation.

The results also demonstrate that the observation time distribution may significantly affect the performance of the CI methods. In our simulations, three observation time distributions are used (Figure 1-a): exp(1), unif(0,2) and gamma(3,1/3). Exp(1) has a monotone decreasing density function, which has the greatest density near 0. Unif(0,2) has a constant density across the limited range of S. Gamma(3,1/3) has the greatest density in the center of the distribution, with the density decreasing in the two tails. In general, the coverage rate is closer to the nominal level at time points where the density of the observation time is great. In contrast, the coverage rate may be poor when this density is very small. For example, when S ~gamma(3,1/3), T ~ $W(0.75, 0.840)$, and at $G(s_0)$ = 0.2 (s_0=0.114), all methods fail to attain 95% coverage rate. This poor performance results from the fact that there are few data points near s_0 for constructing

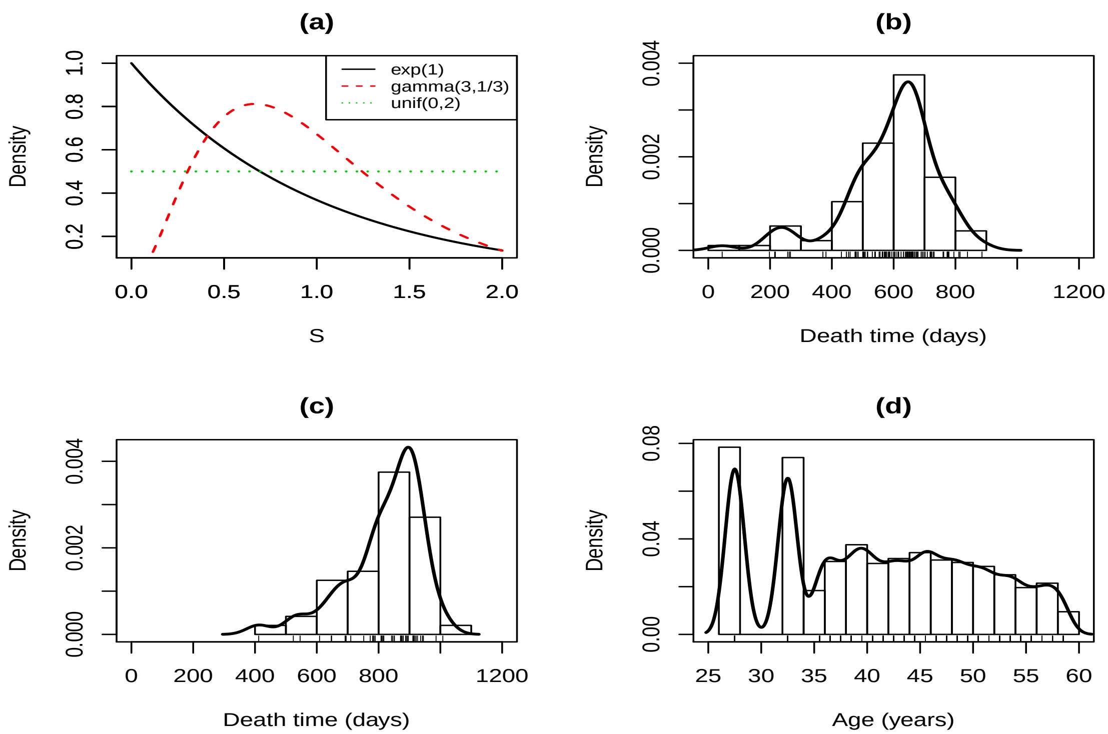

Figure 5.1: (a): Three observation time distributions (exp(1),gamma(3,1/3) and unif(0,2)) used in simulations. (b) and (c): Mice data, kernel density estimates for observed death time in conventional environment and germ-free environment. (d): Menopause data, kernel density estimates for observed age in years

a reliable confidence interval. The density of gamma(3,1/3) at s_0(=0.114) is 0.124, and hence the data is very sparse near this observation time. When the observation time density is low and the sample size is small or moderate, both the nonparametric bootstrap and the bootstrap-Wald CIs outperformed the simple and transformed Wald CIs. Moreover, in that case, the bootstrap-Wald CI is also much narrower than the other CI methods.

5.4 Real examples

5.4.1 Mice lung tumor data

The methods described in this article are applied to mice lung tumor data (Hoel and Walburg 1972). The data consists of 144 male mice from two groups maintained in either conventional environment (96 mice) and germ-free environment (48 mice). The time to tumor onset (in days)

Table 5.1: Biases, estimated standard errors of $\hat{G}(s_0)$, average lengths of confidence intervals and coverage rates of seven types of confidence intervals for the simulation setting (S ~exp(1), T ~exp(1)): non-transformed CIs, logit transformed CIs, log(-log) transformed CIs, nonparametric bootstrap CIs, bootstrap-Wald CIs, logit transformed bootstrap-Wald CIs, and log(-log) transformed bootstrap-Wald CIs at P_0 = $G(s_0)$=0.2, 0.5 and 0.8

				Length						Coverage rate					
P_0	N	Bias	$\hat{sd}(\hat{G})$	wald	logit	llog	nbt	bwald	blogit	wald	logit	llog	nbt	bwald	blogit
	25	-0.036	0.146	0.570	0.576	0.543	0.562	0.443	0.474	0.525	0.517	0.537	0.790	0.727	0.484
	50	-0.034	0.111	0.432	0.466	0.425	0.439	0.383	0.380	0.702	0.692	0.711	0.845	0.769	0.650
	75	-0.026	0.096	0.374	0.409	0.372	0.388	0.357	0.348	0.797	0.810	0.825	0.882	0.812	0.750
	100	-0.026	0.087	0.337	0.370	0.336	0.355	0.333	0.329	0.824	0.867	0.882	0.895	0.838	0.802
.2	200	-0.014	0.071	0.277	0.293	0.273	0.284	0.278	0.301	0.885	0.963	0.955	0.924	0.881	0.903
	500	-0.006	0.055	0.213	0.216	0.209	0.206	0.208	0.229	0.925	0.967	0.956	0.946	0.905	0.947
	800	-0.004	0.047	0.184	0.185	0.181	0.175	0.177	0.189	0.936	0.965	0.959	0.955	0.913	0.941
	1000	-0.004	0.044	0.171	0.172	0.168	0.163	0.164	0.173	0.940	0.965	0.958	0.959	0.919	0.946
	3000	-0.002	0.030	0.118	0.119	0.118	0.112	0.113	0.115	0.948	0.957	0.957	0.967	0.925	0.933
	25	0.028	0.153	0.596	0.569	0.589	0.686	0.676	0.568	0.819	0.917	0.894	0.761	0.816	0.850
	50	0.011	0.130	0.507	0.478	0.491	0.562	0.555	0.528	0.878	0.964	0.936	0.947	0.897	0.936
	75	0.008	0.116	0.453	0.427	0.436	0.478	0.479	0.469	0.898	0.959	0.942	0.963	0.907	0.948
	100	0.004	0.107	0.416	0.395	0.402	0.427	0.431	0.423	0.914	0.953	0.945	0.965	0.914	0.944
.5	200	0.004	0.085	0.332	0.320	0.324	0.331	0.334	0.328	0.930	0.955	0.949	0.969	0.921	0.935
	500	0.002	0.063	0.245	0.240	0.242	0.239	0.241	0.238	0.947	0.959	0.954	0.971	0.923	0.932
	800	0.002	0.054	0.211	0.208	0.209	0.203	0.205	0.203	0.949	0.956	0.954	0.975	0.920	0.924
	1000	0.000	0.050	0.195	0.193	0.194	0.188	0.189	0.188	0.953	0.959	0.957	0.973	0.923	0.929
	3000	0.000	0.035	0.137	0.137	0.137	0.130	0.131	0.130	0.954	0.958	0.957	0.971	0.924	0.926
	25	0.091	0.131	0.511	0.530	0.581	0.525	0.387	0.436	0.358	0.360	0.349	0.437	0.403	0.367
	50	0.065	0.105	0.407	0.448	0.490	0.427	0.432	0.356	0.570	0.579	0.574	0.620	0.607	0.571
	75	0.046	0.089	0.348	0.389	0.421	0.375	0.384	0.329	0.689	0.710	0.705	0.742	0.720	0.685
	100	0.035	0.081	0.314	0.352	0.379	0.344	0.348	0.317	0.747	0.783	0.778	0.806	0.778	0.757
.8	200	0.022	0.064	0.247	0.272	0.287	0.276	0.275	0.295	0.809	0.880	0.877	0.888	0.855	0.891
	500	0.008	0.049	0.190	0.195	0.200	0.204	0.206	0.231	0.851	0.889	0.888	0.942	0.899	0.947
	800	0.006	0.043	0.167	0.169	0.172	0.174	0.176	0.189	0.869	0.893	0.892	0.953	0.902	0.941
	1000	0.005	0.040	0.157	0.158	0.160	0.161	0.163	0.172	0.884	0.905	0.904	0.953	0.909	0.937
	3000	0.002	0.029	0.114	0.114	0.115	0.112	0.113	0.115	0.936	0.946	0.947	0.965	0.919	0.933

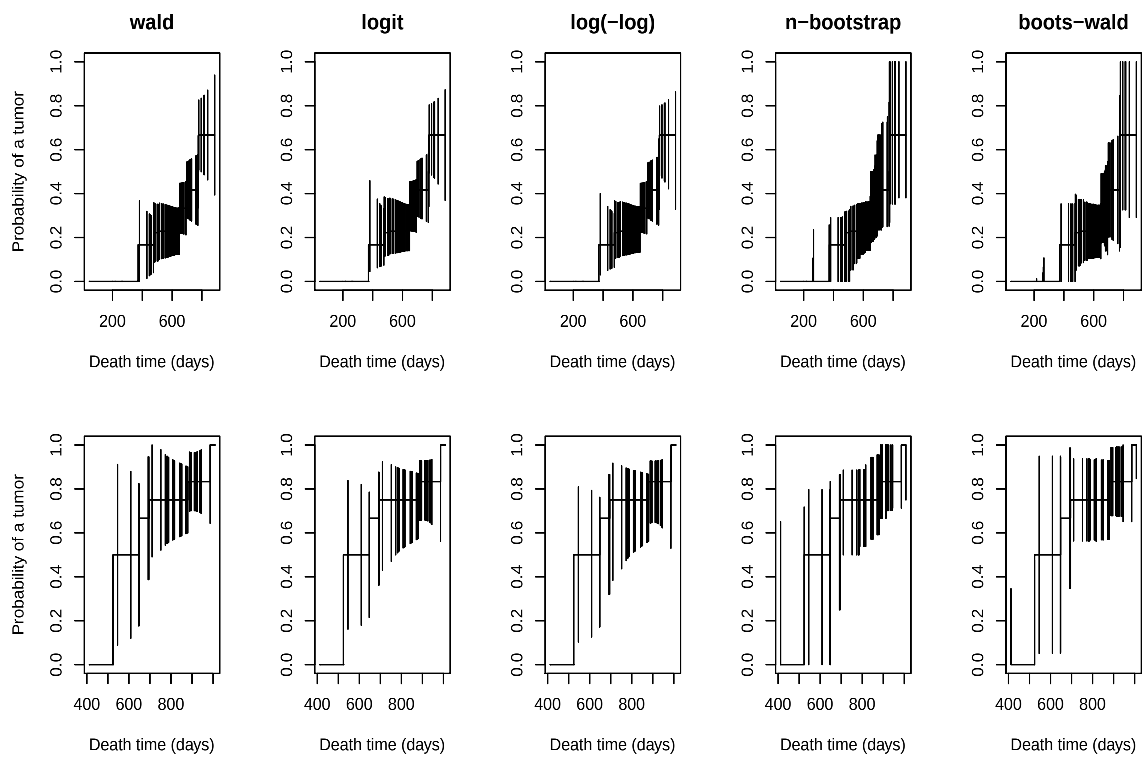

Figure 5.2: Mice lung tumor data. Four types of confidence intervals for conventional environment (the first row) and germfree environment (the second row). From the left, the Wald CIs, the logit-transformed Wald CIs, the log(-log) transformed Wald CIs, the nonparametric bootstrap CIs and the bootstrap-Wald CIs. The estimates of NPMLE for the distribution function of time to lung tumor onset are plotted at observed death times and the corresponding confidence intervals are plotted with vertical lines.

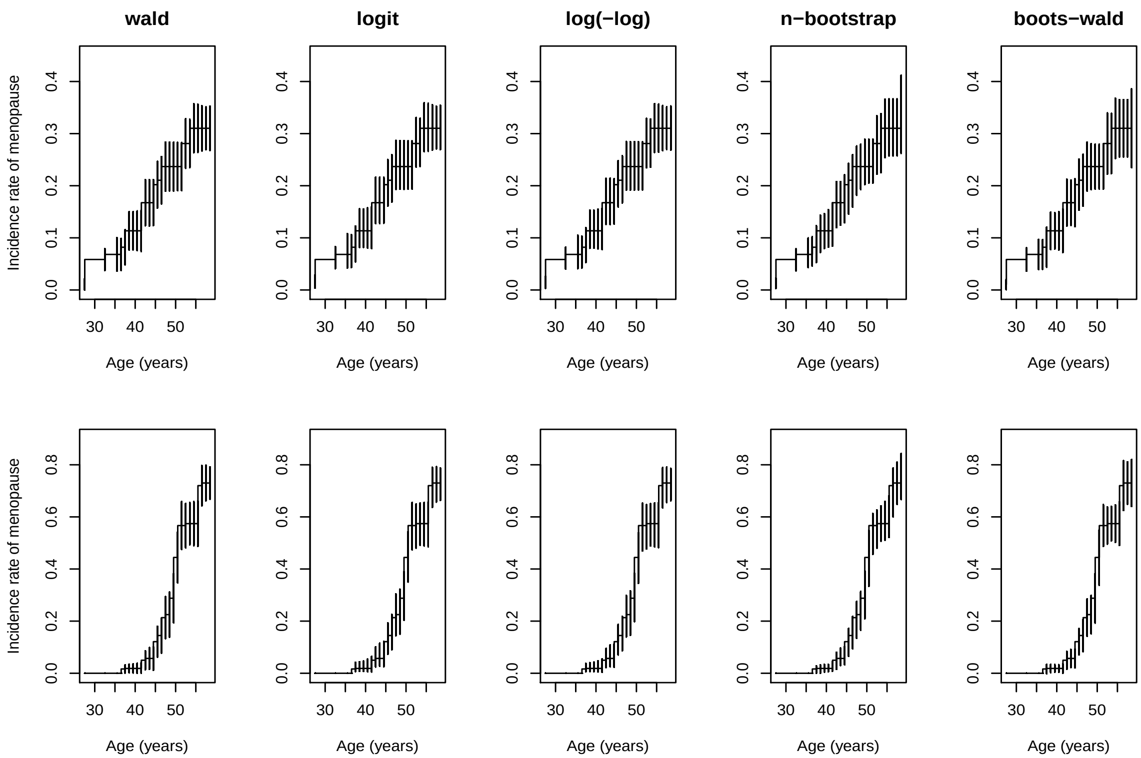

Figure 5.3: Menopause data. Four types of confidence intervals for operative menopause (the first row) and natural menopause (the second row). From the left, the Wald CIs, the logit-transformed Wald CIs, the log(-log) transformed Wald CIs, the nonparametric bootstrap CIs and the bootstrap-Wald CIs. The estimates of NPMLE for the cumulative incidence rate of menopause are plotted at observed ages and the corresponding confidence intervals are plotted with vertical lines.

is of interest, but is not directly observable. Instead, the sacrifice time is available for each mouse and we can observe whether the mouse has a tumor or not at the time of sacrifice.

A 95% CI for the distribution function of time to lung tumour onset (in days) at each observed sacrifice time was constructed using four different methods: the three Wald-type methods and the nonparametric bootstrap, separately for each group. Figure 5.2 shows the NPMLE of distribution function of time to lung tumor onset with the 95% CIs. As with the simulation results for small sample sizes, the transformed Wald CIs seems to be more stable and have shorter lengths than do the non-transformed Wald CIs. The Wald-type CI and bootstrap CI seem to have a similar pattern over time. However, there are some differences. The Wald-type methods shrink to a point when the estimate for the distribution function is zero. In contrast, the bootstrap CI can be constructed at all data points including the points where the estimate for the distribution function is zero. The largest differences between the Wald-type CIs and bootstrap CI occur in the tails where data are sparse. At such time points, the bootstrap CI is highly variable and wider than those based on the Wald-type methods.

By 800 days, the probability of a tumour has leveled off at roughly 0.8 in the germ free environment and 0.7 in the conventional environment. One notes that the onset of tumour occurs earlier in the germ free environment than in the conventional environment, with corresponding probabilities of roughly .5 and .2 at 400 days. One may construct confidence intervals for this difference, which would not be statistically significant, owing to the relatively large standard errors for the respective point estimates, particularly in the germ free group.

To study the reliability of the confidence intervals for this dataset, we used the approach described at the end of Section 2. To repeatedly generate data, we used a kernel smooth estimate of the observation time distribution in each group (see Figure 1(b) and (c)) and fit parametric Weibull models (Keiding et al. 1996) to the data, giving $W(2.04, 1038.07)$ and $W(2.01, 705.72)$ for the conventional and the germ-free environments. The two groups have similar shape parameters with 2.04 and 2.01, but scale parameter is much greater in the conventional environment than in the germ-free environment. Because of the small sample sizes, the performance of the confidence intervals depends heavily on s_0 (see Table 5.10 and 5.11 in the appendix). One sees that the coverage is better at those time points having high density in Figure 1(b) and 1(c). Simulations using alternative models for the failure time distribution gave similar results.

5.4.2 Menopause data

Next we consider menopause data on 2423 females whose age range from 25-29 years from the Health Examination Survey of the National Center for Health Statistics (Macmahon and Worcestor 1966). The females provided their age and menopausal status. For those who had experienced menopause, the causes of menopause, which could be operative or natural, were investigated and reported. Thus the data can be viewed as current status data with competing risk, where the observation time is age and the outcome is menopause with the two competing risks, operative or natural menopause (Maathuis and Hudgens 2011, Jewell et al. 2003).

Jewell et al. (2003) introduced the naive approach to calculate the NPMLE for current status data with competing risks. The approach is to calculate the separate NPMLE for the two competing risks by applying isotonic regression independently, after censoring individuals who experience competing events. We followed this approach to construct CIs for the two menopause groups, which is equivalent to the Wald-type methods we consider in this article. Figure 5.3 displays the confidence intervals for the cumulative incidence functions of operative and natural menopause computed via the Wald-based methods and bootstrap. As in simulation results for large sample sample sizes, the four types of CIs are quite similar, with the intervals being substantially narrower and more informative than those in the mice tumour example.

One observes that operative menopause occurs at a relatively constant rate between ages 30 and 60, while natural menopause only initiates after age 40 and accelerates quickly after age 45. In fact, the two curves cross between ages 45 and 50. By age 60, all women have experienced either operative or natural menopause, with the corresponding difference in their respective probabilities, 0.3 and 0.7, appearing to be highly significant.

Employing the reliability assessment described previously, we simulated using the kernel smooth estimate of the observation time distribution which is the same for both types of menopause and parametric fits to the cumulative incidence functions using the naive likelihood approach Jewell et al. (2003), which yielded $W(10.73, 55.07)$ and $W(3.67, 71.83)$ for the natural menopause and the operative menopause. In contrast to the mice data, all of the confidence intervals at the examined time points are reliable and not sensitive to s_0 because of large sample size (Table 5.12 and 5.13 in the Appendix). The coverage rate of the Wald CI varies from 0.938 to 0.948 for the natural menopause, and from 0.934 to 0.944 for the operative menopause.

5.5 Conclusion

We introduced the transformation of Wald-based CI in binary isotonic regression. Our extensive simulations show that the transformed Wald-CIs have a competitive performance compared to LRT-based CI developed by Banerjee and Wellner (2005) and outperform the non-transformed Wald-CI in small and moderate samples, where improvements are achieved in the coverage rates across a wide range of observation time and failure time scenarios.

We also studied the nonparametric bootstrap CI based on the NPMLE. Although there is no theoretical justification for bootstrap CI for isotonic regression, our simulation shows that bootstrap seems to perform reasonably well in many settings in the sense that it attains a nominal coverage rate, 95%, even in large sample size setting. This is also supported by menopause data whose sample size is large, in which bootstrap CI is quite similar to the Wald type CIs. We also introduced the bootstrap transformed and untransformed Wald CI, where the constant C is estimated by the bootstrap instead of by kernel smoothing. In simulations, they performed reasonably well for small and moderate sample sizes. However, in some cases, they did not attain a nominal coverage rate for large sample sizes, similar to the nonparametric bootstrap. Both the nonparametric bootstrap and the bootstrap-Wald CIs outperformed the simple and transformed Wald-type CIs for the data where the observation time density is low and the sample size is small or moderate.

Our study showed that the performances of confidence interval methods are affected by the observation time (s_0) where a confidence interval is calculated. Near the left and right tails (or boundaries) of the observation time distribution, the coverage rate of CI tends to be lower than in the center of the distribution. The shape of the distributions of the observation and failure time also significantly contribute to the difference in the performance of confidence interval methods.

To give a rough idea as to the reliability of the confidence intervals in a particular dataset, we proposed a simple ad hoc method based on repeatedly simulating under an assumed model for the observation and failure times. In practice, the simulation model may be estimated using the observed data, either using nonparametric or parametric models. The real data examples showed how this technique may be useful. The results show that for the mice data the confidence intervals may perform poorly, with coverage varying substantially across s_0. For the menopause data all of the confidence intervals are reliable because of large sample size.

Transformed Wald confidence intervals for differences between nonparametric estimates based on two independent samples, as in the mice tumour study, may be constructed in a similar fashion to those from a single sample. Here, one can show that the distribution of the difference of two independent NPMLEs is again a scaled version of Z, where the scale factor for the difference involves only the two scale factors for the individual estimates. The distribution of the transformed difference may then be obtained via the Delta method, as in Section 2. The resulting inferences require only the NPMLEs and the estimated scale factors in each group.

The application of the confidence interval methodology to the menopause data in Section 4.2 is the first instance in which theoretically justified confidence intervals have been utilized in nonparametric estimation of the cumulative incidence function with current status competing risks data via naive estimation (Jewell et al. 2003). The construction of such intervals based on full maximum likelihood is unclear, owing to the complicated limiting distribution of the resulting estimators, which is not a simple scaling of Z. Further work is needed to determine whether practicable Wald intervals may be constructed using NPMLE from full maximum likelihood in the competing risks set-up.

BIBLIOGRAPHY

Abadie, A. (2003), "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, 113, 231–63.

Akritas, M. G., Murphy, S. A., and LaValley, M. P. (1995), "The Theil-Sen estimator with doubly censored data and applications to astronomy," *Journal of the American Statistical Association*, 90, 170–177.

Amemiya, T. (1974), "The nonlinear two-stage least-squares estimator," *Journal of Econometrics*, 2, 105–10.

— (1982), "Two-stage least absolute deviations estimators," *Econometrica*, 50, 687–711.

Andre, T., Boni, C., Mounedji-Boudiaf, L., Navarro, M., Tabernero, J., Hickish, T., Topham, C., Zaninelli, M., Clingan, P., Bridgewater, J., Tabah-Fisch, I., and De Gramont, A. (2004), "Oxaliplatin, uorouracil, and leucovorin as adjuvant treatment for colon cancer." *New England Journal of Medicine*, 350, 2343–2351.

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, 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, 929–934.

Banerjee, M. and Wellner, J. A. (2005), "Confidence intervals for Current Status Data." *Scandinavian Journal of Statistics*, 32, 405–424.

Bernhardt, P. W., Wang, H. J., and Zhang, D. (2014), "Flexible modeling of survival data with covaraites subject to detection limits via multiple imputation," *Computational Statistics and Data Analysis*, 69, 81–91.

Betensky, R. A., Rabinowitz, D., and Tsiatis, A. A. (2011), "Computatonally simple accelerated failure time regression for interval censored data," *Biometrika*, 88, 703–711.

Bie, O., Borgan, O., and Liestol, K. (1987), "Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties." *Scandinavian Journal of Statistics*, 14, 221–233.

Bijwaard, G. E. (2008), "Instrumental variable estimation for duration data," *Working paper, Tinbergen Institute.*

Bollen, K. A. (1996), "An alternative two stage least squares (2SLS) sstimator for latent variable equations." *Psychometrika*, 61, 109–121.

Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., and Chen, F. (2007), "Latent variable models under misspecification: Two-Stage Least Squares (2SLS) and Maximum Likelihood (ML) Estimators." *Social Science Research*, 36, 48–86.

Borgan, O. and Liestol, K. (1990), “A note on confidence intervals and bands for the survival function based on transformations.” *Scandinavian Journal of Statistics*, 17, 35–41.

Brannas, K. (2000), “Estimation in a duration model for evaluating educational programs,” *IZA Discussion paper*.

Brookhart, M. A., Rassen, J. A., and Schneeweiss, S. (2010), “Instrumental variable methods in comparative safety and effectiveness research,” *Pharmacoepidemiology and Drug Safety.*, 19, 537–554.

Brookhart, M. A., Wang, P. S., Solomon, D. H., and Schneeweiss, S. (2006), “Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable.” *Epidemiology.*, 17, 268–275.

Brumback, B. A., A., H. M., Haneuse, S. J. P. A., and Robins, J. M. (2004), “Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures,” *Statistics in Medicine*, 23, 749–767.

Buckley, J. and James, I. (1979), “Linear regression with censored data.” *Biometrika*, 65, 429–436.

Chen, L. A. and Portnoy, S. (1996), “Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equations models,” *Communications in Statistics - Theory and Methods*, 25, 1005–1032.

Choi, B. Y., Fine, J. P., and Brookhart, M. A. (2014), “Two-stage estimation of structural instrumental variable models with coarsened data,” *Biometrika* (submitted).

Flanders, W. D., Klein, M., Darrow, L. A., Strickland, M. J., Sarnat, S. E., Sarnat, J. A., Waller, L. A., Winquist, A., and Tolbert, P. E. (2011), “A method for detection of residual confounding in time-series and other observational studies,” *Epidemiology*, 22, 59–67.

Fygenson, M. and Ritov, Y. (1994), “Monotone estimating equations for censored data,” *Annals of Statistics*, 22, 732–746.

Ghosh, D., Banerjee, M., and Biswas, P. (2008), “Inference for constrained estimation of tumor size distributions,” *Biometrics*, 64, 1009–1017.

Gonzalez-Manteiga, W. and Cadarso-Suarez, C. (1994), “Asymptotic properties of a generalized Kaplan-Meier estimatoar with some applications,” *Journal of Nonparametric Statistics*, 4, 65–78.

Groeneboom, P. and Wellner, J. A. (1992), *Information Bounds and Nonparametric Maximum Likelihood Estimation.*, Boston: Birkhäuser.

— (2001), “Computing Chernoff’s distribution.” *Journal of Computational and Graphical Statistics*, 10, 388–400.

Gustafson, P., McCandless, A. R., and Richardson, S. (2010), “Simplified Bayesian sensitivity analysis for mismeasured and unobserved confounders,” *Biometrics*, 66, 1129–1137.

Hansen, L. B. (1982), "Large sample properties of generalized method of moment estimators," *Econometrica*, 46, 1029–1054.

Heckman, J. J. (1978), "Dummy endogeneous variables in a simultaneous equation system," *Econometrica*, 46, 931–959.

Heitjan, D. F. and Rubin, D. B. (1990), "Inference from Coarse data via multiple imputation with application to age heaping," *Journal of the American Statistical Association*, 85, 304–314.

— (1991), "Ignorability and coarse data." *Annals of Statistics*, 19, 2244–2253.

Hernan, M. A. and Robins, J. M. (2006), "Instruments for causal inference: An epidemilogist's dream ?" *Epidemiology*, 17, 360–372.

Hoel, D. G. and Walburg, H. E. (1972), "Statistical analysis of survival experiments." *Journal of National Cancer Institute*, 49, 361–372.

Huang, L. and Jin, Z. (2006), *lss: the accelerated failure time model to right censored data based on least-squares principle.*, R package version 0.52.

Jackson, L. A., Jackson, M. L., Nelson, J. C., Neuzil, K. M., and Weiss, N. S. (2006), "Evidence of bias in estimates of influenza vaccine effectiveness in seniors," *International Journal of Epidemiology*, 35, 337–344.

Jewell, N. P., Van Der Laan, M., and Henneman, T. (2003), "Nonparametric estimation from current status data with competing risks." *Biometrika*, 90, 183–197.

Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003), "Rank-based inference for the accelerated failure time model." *Biometrika*, 90, 341–353.

Jin, Z., Lin, D. Y., and Ying, Z. (2006a), "Rank regression analysis of multivariate failure time data based on marginal linear models." *Scandinavian Journal of Statistics*, 33, 1–23.

— (2006b), "On least squares regression with censored data," *Biometrika*, 93, 147–161.

Jin, Z., Ying, Z., and Wei, L. J. (2001), "A simple resampling method by perturbing the minimand." *Biometrika*, 88, 381–390.

Johnson, B. A. and Tsiatis, A. A. (2004), "Estimationg mean response as a function of treatment duration in an observatinal study, where duration may be informatively censored," *Biometrics*, 60, 315–323.

Keiding, N., Begtrup, K., Scheike, T. H., and Hasibeder, G. (1996), "Estimation from current-status data in continuous time." *Lifetime Data Analysis*, 2, 119–129.

Lai, T. L. and Ying, Z. (1991aa), "Large sample theory of a modified Buckley-James estimator for regression analysis with censored data," *Annals of Statistics*, 19, 1370–1402.

— (1991ab), "Rank regression methods for left-truncated and rigth-censored data," *Annals of Statistics*, 19, 531–556.

Lin, D. Y., Psaty, B. M., and Kronmal, R. A. (1998), "Assessing the sensitivity of regression results to unmeasured confounders in observational studies," *Biometrics*, 54, 948–963.

Lipsitch, M., Tchetgen, E. T., and Cohen, T. (2010), "Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies," *Epidemiology*, 21, 383–388.

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, 100–105.

Loeys, T., Goetghebeur, E., and Vandebosch, A. (2005), "Causal proportional hazards models and time-constant exposure in randomized clinical trials." *Lifetime Data Analysis*, 11, 435–449.

Lumley, T. and Sheppard, L. (2000), "Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control analyses," *Environmetrics*, 11, 705–717.

Maathuis, M. H. and Hudgens, M. G. (2011), "Nonparametric inference for competing riks current status data with continuous, discrete or grouped observation times." *Biometrika*, 98, 325–340.

Mack, C. D., Brookhart, M. A., Glynn, R. J., Meyer, A. M., Carpenter, W. R., Sandler, R. S., and Sturmer, T. (2015), "Comparative effectiveness of oxaliplatin vs. 5-flourouricil in older adults: An instrumental variable analysis." *Epidemiology*, 26, 690–699.

Macmahon, B. and Worcestor, J. (1966), "Age at menopause, United States 1960–1962." *National Center for Health Statistics. Vital Health Statistics*, 11, 161–173.

McClellan, M., McNeil, B. J., and Newhouse, J. P. (1994), "Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables," *The Journal of the American Medical Associatoin*, 272, 859–866.

Newey, W. K. (1990), "Efficient instrumental variables estimation of nonlinear models," *Econometrica*, 58, 809–837.

Newey, W. K. and McFadden, D. (1994), *Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics*, Elsevier Science B.V., Amsterdam, pp. 2111-2245, Chapter 36.

Newey, W. K. and Powell, J. L. (2003), "Instrumental variable estimation of nonparametric models," *Econometrica*, 71, 1565–1578.

Newey, W. K., Powell, J. L., and Vella, F. (1999), "Nonparametric estimation of triangular simultaneous equation models," *Epidemiology*, 67, 565–603.

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

Pan, W. (2000), “Smooth estimation of the survival function for interval censored data.” *Statistics in Medicine*, 19, 2611–2624.

Pang, L., Lu, W., and Wang, H. J. (2014), “Local Buckley James estimation for heteroscedastic accelerated failure time model,” *Statistica Sinica* (in press).

Pearl, J. (2002), “Causal inference in the health sciences: A conceptual introduction,” *Health Services and Outcomes Research Methodology*, 2, 189–220.

Pollard, D. (1990), *Epirical processes: Theory and Application.*, Hayward, CA: IMS.

Powell, J. L. (1984), “Least absolute deviations for the censored regression models,” *Journal of Econometrics*, 25, 303–325.

— (1986a), “Symmetrically trimmed least squares estimation for tobit models,” *Econometrika*, 54, 1435–1460.

— (1986b), “Censored regression quantiles,” *Journal of Econometrics*, 32, 143–155.

Prentice, R. L. (1978), “Linear rank tests with right censored data,” *Biometrika*, 65, 167–179.

Rabinowitz, D., Tsiatis, A. A., and Aragon, J. (1995), “Regression with interval-censroed data,” *Biometrika*, 82, 501–513.

Ritov, Y. (1990), “Estimation in a linear regression model with censored data,” *Annals of Statistics*, 18, 303–328.

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, 2609–2631.

Rosenbaum, P. R. and Rubin, D. B. (1983), “Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome,” *Journal of the Royal Statistical Society, Series B*, 45, 212–218.

Schneeweiss, S., Solomon, D. H., Wang, P. S., Rassen, J., and Brookhart, M. A. (2006), “Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective cyclooxygenase 2 inhibitors and nonselective nonsteroidal antiinflammatory drugs: an instrumental variable analysis.” *Arthritis Rheum*, 54, 3390–3398.

Self, S. G. and Grossman, E. A. (1986), “Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers,” *Biometrics*, 42, 521–530.

Sheather, B. J. and Jones, M. C. (1991), “A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation.” *Journal of the Royal Statistical Society, Series B*, 53, 683–690.

Smith, G. D. (2008), “Assessing intrauterine influences on offspring health outcomes: Can epidemiological studies yield robust findings?” *Basic & Clinical Pharmacology & Toxicology*, 102, 245–256, International Conference on Foetal Programming and Development Toxicity, Torshavn, DENMARK, MAY 20-24, 2007.

— (2012), “Negative Control Exposures in Epidemiologic Studies,” *Epidemiology*, 23, 350–351.

Smith, G. D., Lawlor, D. A., Harbord, R., Timpson, N., Rumley, A., Lowe, G. D. O., Day, I. N. M., and Ebrahim, S. (2005), “Assocation of C-reactive protein with blood pressure and hypertension: life course confouding and Mendilan randomization tests of causality,” *Arteriosclerosis, Thrombosis, and Vascular Biology*, 25, 1051–1056.

Staiger, D. and Stock, J. H. (1997), “Instrumental variables regression with weak instruments,” *Econometrica*, 65, 557–586.

Stock, J. and Yogo, M. (2005), *Testing for Weak Instruments in Linear IV Regression*, New York: Cambridge University Press, pp. 80–108.

Stukel, T. A., Fisher, E. S., Wennberg, D. E., Alter, D. A., Gottlieb, D. J., and Vermeulen, M. J. (2007), “Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods.” *The Journal of the American Medical Associatoin*, 297, 278–285.

Tchetgen Tchetgen, E. (2014), “The control outcome calibration approach for causal inference with unobserved confouding,” *American Journal of Epidemiology*, 179, 633–640.

Terza, J. V., Basu, A., and Rathouz, P. J. (2008), “Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling,” *Journal of Health Economics*, 27, 531–543.

Theil, H. (1953), “Repeated least-squares applied to complete equation systems,” *Centraal Planbureau Memorandum*.

Tian, L. and Cai, T. (2006), “On the accelerated failure time model for current status data and interval censored data,” *Biometrika*, 93, 329–342.

Tsiatis, A. A. (1990), “Estimating regression parameters using linear rank tests for censored data,” *Annals of Statistics*, 18, 354–372.

VanderWeele, T. J. (2008), “Sensitivity analysis: distributional assumptions and confouding assumptions,” *Biometrics*, 64, 645–649.

VanderWeele, T. J., Mukherjee, B., and Chen, J. (2012), “Sensitivity analysis for interactions under unmeasured confounding,” *Statistics in Medicine*, 31, 2552–2564.

Wang, H. J. and Feng, X. (2012), “Multiple impuation for M-regression with censored covariates,” *Journal of the American Statistical Association*, 107, 194–204.

White, H. (1980), “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica*, 48, 817–838.

— (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.

Ying, Z. (1993), "A large sample study of rank estimation for censored regression data," *Annals of Statistics*, 21, 76–99.