On 2-Step Residual Inclusion Estimator for Instrument Variable Additive Hazards Model

Binyan Jiang

Department of Applied Mathematics

The Hong Kong Polytechnic University

Jialiang Li*

Department of Statistics and Applied Probability

National University of Singapore

Duke-NUS Graduate Medical School

Singapore Eye Research Institute

and

Jason Fine

Department of Biostatistics

Department of Statistics and Operations Research

University of North Carolina, Chapel Hill

Abstract

Instrumental variable (IV) methods are popular in non-experimental settings to estimate the causal effects of scientific interventions. These approaches allow for the consistent estimation of treatment effects even if major confounders are unavailable. There have been some extensions of IV methods to survival analysis recently. We specifically consider the 2-Step Residual Inclusion (2SRI) estimator for the additive hazards model in this paper. Assuming linear structural equation models for the hazard function, we may attain a closed-form, two-stage estimator for the causal effect in the additive hazards model. The asymptotic properties of the estimators are rigorously established and the resulting inferences are shown to perform well in simulation studies.

Keywords: Additive hazards model; instrumental variable; two-stage least squares estimation; survival analysis.

1 Introduction

Instrumental variable (IV) approaches have been widely used in Econometric studies and recently became an active research area for biostatistical community. One of the areas that still need our attention is how to apply IV approaches properly for censored survival data. Li et al. (2015) first proposed to adapt the familiar 2-step regression approaches for additive hazards model. While other survival models may also be considered in practice, it remains unclear whether the standard IV approaches

*email: stalj@nus.edu.sg

can be easily accommodated. We confine our attention to additive hazards model in this paper.

Most applications of IV methods in the medical and epidemiologic literature are based on two-stage least squares estimation of structural equation models. Robins and Tsiatis (1991) proposed an IV estimator to correct for non-compliance in randomized trials with the error distributions in the structural equations unspecified. These methods require that the censoring time is always observed and are not applicable with random loss to follow-up where only the minimum of the event and censoring times is known. Tang and Lee (1998) proposed parametric maximum likelihood estimators, with numerical studies presented in Posner and Baker (2000). Related work on parametric discrete time models is found in Muthen and Masyn (2005) and Chen et al. (2011). For the usual right censoring set-up, most of the above developments rely heavily on specified parametric models and may not be attractive in the medical study where semiparametric models with unknown error distributions are the default. Loeys and Goetghebeur (2003) extended the Robins-Tsiatis approach to causal proportional hazards models with observed censoring times. Bijwaard (2008) studied endogeneity for duration data under the generalized accelerated failure time model and extended the IV approach to rank estimation with known censoring times. Terza et al. (2008) considered the two-stage predictor substitution as well as the two-stage residual inclusion for parametric likelihood estimation with uncensored duration data. Li et al. (2015) proposed a new methodology to address the usual right censoring encountered in medical applications. Moreover, the semiparametric additive hazards model considered in Li et al. (2015) is more flexible than the parametric models in the previous literature.

Recently Chan (2015) pointed out an alternative way of fitting instrumental variable additive hazards model using the well-known 2-stage residual inclusion (2SRI) approach. This approach complements the 2-stage predictor substitution (2SPS) approach proposed in Li et al. (2015) since 2SPS requires a relatively stronger conditional independence assumption for the random censoring time. In linear regression these two approaches are often interchangeable since they agree numerically and theoretically. For nonlinear models they are in general not identical and thus should be chosen by the practitioners with proper consideration.

Chan (2015) carried out extensive simulation studies to examine the performance of 2SRI approach and compare with 2SPS approach. However, no theoretical works were provided to justify the validity of 2SRI estimator. In this paper we attempt to fill in this gap by carefully establishing the technical conditions and supplying the mathematical argument for the asymptotic properties of 2SRI estimator. Consistency and asymptotic normality for the 2SRI estimator under additive hazards model are given in the main theorems of this paper. In addition, we also carry out an efficiency comparison between 2SRI and 2SPS estimators.

To complement the numerical works in Chan (2015), we select settings where assumptions for 2SRI and 2SPS might not hold. The robustness of the approaches against model misspecification is assessed via our simulations.

2 Estimation

Suppose the true model for explaining the causal effects of covariates on the hazard function of the survival time T is

$$h(t; X_e, \mathbf{X}_o, \mathbf{X}_u) = h_0(t) + \beta_e X_e + \boldsymbol{\beta}_o^T \mathbf{X}_o + \eta, \tag{1}$$

where X_e is an endogenous variable whose coefficient is of interest, $\mathbf{X}_o$ is a $p-$vector of observed exogenous variables, and the error η involves some unobserved covariates $\mathbf{X}_u$. In the literature of lifetime data analysis model (1) is referred to as the additive hazards model (Lin and Ying (1994)).

Further suppose we have an instrumental variable X_I such that

$$X_e = \alpha_c + \alpha_I X_I + \boldsymbol{\alpha}_o^T \mathbf{X}_o + v, \tag{2}$$

where $\alpha_c, \alpha_I, \boldsymbol{\alpha}_o$ are coefficients for the constant intercept, the instrumental variable, and the observed exogenous variable, respectively. The error v may involve unobserved covariates. The model (2) assumes a standard structure for the endogenous variable in IV estimation.

When survival time is subject to right censoring, the observed data consist of n independent realizations of $Y = \min(T, C), \Delta = I(T \leq C)$ and covariates $\{X_e, \mathbf{X}_o, X_I\}$, where C is a random censoring time which may prevent observation of T. We shall denote these n samples as $Y_i = \min(T_i, C_i), \Delta_i = I(T_i \leq C_i)$ and $\{X_{ei}, \mathbf{X}_{oi}, X_{Ii}\}, i = 1, \ldots, n$.

To implement the 2SPS approach, we first obtain $(\hat{\alpha}_c, \hat{\alpha}_I, \hat{\boldsymbol{\alpha}}_o^T)$ with a least squares approach and predict the endogenous variable $\hat{X}_e = \hat{\alpha}_c + \hat{\alpha}_I X_I + \hat{\boldsymbol{\alpha}}_o^T \mathbf{X}_o$. Then we fit the additive hazards model using $\hat{X}_e$ and $\mathbf{X}_o$ as the covariates. The asymptotic properties of the resulting 2SPS estimator were studied in Li et al. (2015).

To implement the 2SRI approach, we first obtain $(\hat{\alpha}_c, \hat{\alpha}_I, \hat{\boldsymbol{\alpha}}_o^T)$ from a first stage least squares estimation and predict the residuals $\hat{v} = X_e - \hat{\alpha}_c - \hat{\alpha}_I X_I - \hat{\boldsymbol{\alpha}}_o^T \mathbf{X}_o$. In the second stage we fit the additive hazards model using $\hat{X}_e$, $\mathbf{X}_o$ and $\hat{\nu}$ as the covariates. The asymptotic properties of the resulting 2SRI estimator were provided in the following section.

3 Asymptotic Properties

We first introduce some notations. Denote $\boldsymbol{\beta} = (\beta_e, \boldsymbol{\beta}_o^T, \gamma_v)^T$ where γ_v is given in condition (C2), $\mathbf{X}_{IOi} = (1, X_{Ii}, \mathbf{X}_{oi}^T)^T$, $\mathbf{X}_{IO} = [\mathbf{X}_{IO1} \cdots \mathbf{X}_{IOn}]^T$, $\mathbf{X}_e = (X_{e1}, \ldots, X_{en})^T$, $\mathbf{X}_{IOEi} = (1, X_{Ii}, \mathbf{X}_{oi}^T, X_{ei})^T$, $\mathbf{X}_{IOE} = [\mathbf{X}_{IOE1} \cdots \mathbf{X}_{IOEn}]^T$, and $v_i = X_{ei} - \alpha_c - \alpha_I X_{Ii} + \alpha_o^T \mathbf{X}_{oi}$.

The residual of the first-stage least squares prediction can be written as

$$\hat{v}_i = X_{ei} - \mathbf{X}_{IOi}^T [\mathbf{X}_{IO}^T \mathbf{X}_{IO}]^{-1} \mathbf{X}_{IOi} \mathbf{X}_e. \tag{3}$$

Denote

$$\mathbf{X}_{EOV} = \begin{bmatrix} (0, 0, \mathbf{0}_p^T, 1) \\ (\mathbf{0}_p, \mathbf{0}_p, \mathbf{I}_p, \mathbf{0}_p) \\ (-\mathbf{X}_e^T \mathbf{X}_{IO} [\mathbf{X}_{IO}^T \mathbf{X}_{IO}]^{-1}, 1), \end{bmatrix}$$

where $\mathbf{0}_p$ is a vector of zeros with length p and $\mathbf{I}_p$ is the $p \times p$ dimensional identity matrix. Denote the true value of the coefficients as $\boldsymbol{\beta}_T = (\beta_{Te}, \boldsymbol{\beta}_{To}^T, \gamma_{Tv})^T$ and let

$$\bar{\mathbf{X}}_{IOE}(t) = \frac{\sum_{i=1}^n A_i(t)\mathbf{X}_{IOEi}}{\sum_{i=1}^n A_i(t)},$$

where $A_i(t) = I(Y_i \geq t)$ is a 0-1 at-risk process indicating whether the subject i is at risk at time t.

We impose the following technical conditions:

(C1) $X_I, \mathbf{X}_u$ and $\mathbf{X}_o$ are mean zero random variables with bounded supports.

(C2) $\eta = \gamma_v v + \epsilon$. $\alpha_I \neq 0$, $\gamma_v \neq 0$.

(C3) X_I is independent of $(T, \mathbf{X}_u)$ conditional on $\mathbf{X}_o$.

(C4) $\mathbf{X}_u$ and $\mathbf{X}_o$ are independent.

(C5) v is independent of X_I and $\mathbf{X}_o$ with variance $Var(v) = \sigma^2$; ϵ is a white noise and is independent of $X_I, \mathbf{X}_u$ and $\mathbf{X}_o$.

(C6) C is conditionally independent of T given $(X_e, X_I, \mathbf{X}_o)$.

(C7) The matrix $\sum_{i=1}^n \int_0^\infty [\mathbf{X}_{IOE}(\mathbf{X}_{IOi} - \bar{\mathbf{X}}_{IO}(t))]^{\otimes 2} A_i(t)dt$ is non-singular.

(C8) There exists $(p+2) \times (p+2)$ positive definite matrices Σ, Ω and Ψ such that

$$n^{-1}\sigma^2\gamma_{Tv}^2 \sum_{i=1}^n \int_0^\infty [\mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))]\mathbf{X}_{IOi}^T[\mathbf{X}_{IO}^T\mathbf{X}_{IO}]^{-1}\mathbf{X}_{IOi}$$
$$[\mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))]^T A_i(t)dt \stackrel{a.s.}{\to} \Sigma,$$

$$\frac{1}{n}\sum_{i=1}^n \int_0^\infty [\mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))]^{\otimes 2} A_i(t)dt \stackrel{a.s.}{\to} \Omega,$$

and

$$\sum_{i=1}^{n}\int_{0}^{\infty}[\mathbf{X}_{EOV}(\mathbf{X}_{IOEi}-\bar{\mathbf{X}}_{IOE}(t))]^{\otimes 2}dN_i(t)\stackrel{a.s.}{\to}\Psi.$$

Here $N_i(t)=I(Y_i\leq t,\Delta_i=1)$ is the counting process for the ith subject.

Denote the 2SRI estimator of $\boldsymbol{\beta}$ as $\hat{\boldsymbol{\beta}}$. $\hat{\boldsymbol{\beta}}$ is then the solution of $U(\boldsymbol{\beta})=0$ where

$$U(\boldsymbol{\beta})=\sum_{i=1}^{n}\int_{0}^{\infty}[\mathbf{X}_{EOV}(\mathbf{X}_{IOEi}-\bar{\mathbf{X}}_{IOE}(t))][dN_i(t)-\boldsymbol{\beta}^T\mathbf{X}_{EOV}\mathbf{X}_{IOEi}A_i(t)dt]. \quad (4)$$

It can be shown that $\hat{\boldsymbol{\beta}}$ has the following closed-form:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left\{\sum_{i=1}^{n}\int_{0}^{\infty}[\mathbf{X}_{EOV}(\mathbf{X}_{IOEi}-\bar{\mathbf{X}}_{IOE}(t))]^{\otimes 2}A_i(t)dt\right\}^{-1}\\ &\quad\times\left\{\sum_{i=1}^{n}\int_{0}^{\infty}[\mathbf{X}_{EOV}(\mathbf{X}_{IOEi}-\bar{\mathbf{X}}_{IOE}(t))]dN_i(t)\right\}.\end{aligned}$$

We denote the true parameter as $\boldsymbol{\beta}_T$ in the following theorems.

Theorem 3.1 *(Strong Consistency.) Under Conditions (C1) - (C7), we have* $\hat{\boldsymbol{\beta}}\to\boldsymbol{\beta}_T$ *almost surely as* $n\to\infty$.

Theorem 3.2 *(Asymptotic Normality.) Under Conditions (C1) - (C8), we have* $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_T)\to N(0_{p+2},\Omega^{-1}(\Psi+\Sigma)\Omega^{-1})$ *in distribution as* $n\to\infty$.

Following Lin and Ying (1994), the estimator for the confounded (with the unobserved covariates) baseline cumulative hazard $H_0^{\dagger}(t)=-\log S_0^{\dagger}(t)$ can be estimated by

$$\hat{H_0}^{\dagger}(t) = \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - \hat{\boldsymbol{\beta}}^T \mathbf{X}_{EOV}\mathbf{X}_{IOEi}A_i(u)du\}}{\sum_{i=1}^n A_i(u)}.$$

Consequently, the covariate specific survival function $S(t; x_e, \mathbf{x}_o, v)$ as in (5) can be estimated by

$$\hat{S}(t; x_e, \mathbf{x}_o, v) = \exp\{-\hat{H_0}^{\dagger}(t) - t(\hat{\beta}_e x_e + \hat{\boldsymbol{\beta}}_o^T \mathbf{x}_o + \hat{\gamma}_v v)\}.$$

Assuming that:

(C9) For a given $\tau < \infty$, there exists continuous and bounded functions $g(t)$ and $\mathbf{G}(t)$ such that

$$\sup_{t\in(0,\tau]} \left| \int_0^t \frac{n\sum_{i=1}^n dN_i(u)}{\{\sum_{i=1}^n A_i(u)\}^2} - g(t) \right| \xrightarrow{p} 0$$

$$\sup_{t\in(0,\tau]} \left\| \sum_{i=1}^n \mathbf{X}_{EOV}\mathbf{X}_{IOEi} \int_0^t \frac{dN_i(u)}{\sum_{i=1}^n A_i(u)} - \mathbf{G}(t) \right\|_{max} \xrightarrow{p} 0.$$

The following theorem establishes the asymptotic properties of $\hat{S}^{\dagger}(t; x_e, \mathbf{x}_o, v)$:

Theorem 3.3 *(Weak Convergence of Survival Function) Under Conditions (C1) - (C9), we have as $n \to \infty$,*

$$\sup_{t\in(0,\tau]} \left| \hat{S}(t; x_e, \mathbf{x}_o, v) - S(t; x_e, \mathbf{x}_o, v) \right| \xrightarrow{p} 0.$$

Furthermore, $\sqrt{n}\hat{S}(t; x_e, \mathbf{x}_o, v) - S(t; x_e, \mathbf{x}_o, v)$ converges weakly to a mean zero Gaussian process over $t \in (0, \tau]$ and the covariance function at (t_1, t_2) for $t_1, t_2 \in (0, \tau]$ is *given by:*

$$S(t_1; x_e, \mathbf{x}_o, v)S(t_2; x_e, \mathbf{x}_o, v)\{g(\min(t_1, t_2)) +$$

$$(x_e, \mathbf{x}_o^T, v)\Omega^{-1}(\Psi + \sigma^2\Sigma)\Omega^{-1}(x_e, \mathbf{x}_o^T, v)^T t_1 t_2\} + (x_e, \mathbf{x}_o^T, v)\Omega^{-1}[\mathbf{G}(t_1)t_2 + \mathbf{G}(t_2)t_1].$$

We next establish the equivalence of the 2SRI and 2SPS estimators of β_e under appropriate assumptions. Denote $\hat{\beta}_e$ and $\tilde{\beta}_e$ as the estimators of β_e using 2SRI and 2SPS respectively. Write $\hat{X}_{ei} = \mathbf{X}_{IOi}^T[\mathbf{X}_{IO}^T\mathbf{X}_{IO}]^{-1}\mathbf{X}_{IOi}\mathbf{X}_e$, $\mathbf{X}_{si} = (\hat{X}_e, \mathbf{X}_{oi}^T)^T$, $\mathbf{X}_s = [\mathbf{X}_{s1}, \ldots, \mathbf{X}_{sn}]^T$ and denote:

$$\bar{\mathbf{X}}_s(u) = \frac{\sum_{i=1}^n A_i(u)\mathbf{X}_{si}}{\sum_{i=1}^n A_i(u)}.$$

Similarly, write $\mathbf{X}_{svi} = (\hat{X}_e, \mathbf{X}_{oi}^T, \hat{v}_i)^T$ and $\mathbf{X}_{tv} = [\mathbf{X}_{sv1}, \ldots, \mathbf{X}_{svn}]^T$ and define

$$\bar{\mathbf{X}}_{sv}(u) = \frac{\sum_{i=1}^n A_i(u)\mathbf{X}_{sv}}{\sum_{i=1}^n A_i(u)}.$$

Denote

$$\epsilon_{e,n} = \frac{1}{n}\sum_{i=1}^n \int_0^\infty \left(\hat{X}_{ei} - \frac{\sum_{j=1}^n A_j(u)\hat{X}_{ej}}{\sum_{j=1}^n A_j(u)}\right)\left(\hat{v}_i - \frac{\sum_{j=1}^n A_j(u)\hat{v}_j}{\sum_{j=1}^n A_j(u)}\right) A_i(u)du$$

Similarly, denote the jth element of

$$\frac{1}{n}\sum_{i=1}^n \int_0^\infty \left(\mathbf{X}_{oi} - \frac{\sum_{j=1}^n A_j(u)\mathbf{X}_{oj}}{\sum_{j=1}^n A_j(u)}\right)\left(\hat{v}_i - \frac{\sum_{j=1}^n A_j(u)\hat{v}_j}{\sum_{j=1}^n A_j(u)}\right) A_i(u)du$$

as $\epsilon_{oj,n}$, and write

$$
\begin{aligned}
B &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}[\mathbf{X}_{si}-\bar{\mathbf{X}}_s(u)]^{\otimes 2}A_i(u)du,\\
C &= (\epsilon_{e,n},\epsilon_{o1,n},\ldots,\epsilon_{op,n})^T,\\
D &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\left(\hat{v}_i-\frac{\sum_{j=1}^{n}A_j(u)\hat{v}_j}{\sum_{j=1}^{n}A_j(u)}\right)^2 A_i(u)du,\\
\alpha &= (D-C^TB^{-1}C)^{-1},\\
Y_N &= (Y_{N,1},\ldots,Y_{N,p+2})=\frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}[\mathbf{X}_{svi}-\bar{\mathbf{X}}_{sv}(u)]dN_i(u)
\end{aligned}
$$

Assume that the conditions for 2SPS Li et al. (2015) and conditions (C1)-(C8) hold. The following theorem indicates that when η in linear in $\mathbf{X}_u$ as in 2SRI, the two estimators obtained using 2SRI and 2SPS are asymptotically equivalent. For any given matrix M we denote its first row as $M_{1,\cdot}$. Similarly, the first element of a vecot V is denoted as V_1.

Theorem 3.4 *Uder the assumptions for 2SPS and conditions (C1)-(C9), we have,*

$$\hat{\beta}_e=\tilde{\beta}_e+\alpha[B^{-1}C(B^{-1}C)^T]_{1,\cdot}(Y_{N,1},\ldots,Y_{N,p+1})^T-\alpha(B^{-1}C)_1Y_{N,p+2}=\tilde{\beta}_e+O_p(n^{-1/2}).$$

Consequently we have,

$$
\begin{aligned}
Var(\hat{\beta}_e) &= Var(\tilde{\beta}_e)+Var(\alpha(B^{-1}C)_1Y_{N,p+2})-2Cov(\tilde{\beta}_e,\alpha(B^{-1}C)_1Y_{N,p+2})+O(n^{-3/2})\\
&= Var(\tilde{\beta}_e)+O(n^{-1}).
\end{aligned}
$$

4 Numerical study

Chan (2015) clearly indicated a potential weakness of 2SRI in that the first stage residual η and the second stage residual ν in the two stage models must be linearly associated (Condition (C2) above). Such a structural requirement could be quite restrictive in practice. For example, using notations from Li et al. (2015), let us consider a very common setting where $\eta = \boldsymbol{\beta}_u^T \mathbf{X}_u$ and $\nu = \boldsymbol{\alpha}_u^T \mathbf{X}_u + \epsilon$. Here $\mathbf{X}_u$ denotes the $q-$vector of unobserved confounding variables. The condition (C3) can be trivially satisfied if $q = 1$. Sometimes it is reasonable to assume the unobserved confounding effects could be attributed to a single latent variable. This kind of specification has been practiced widely in longitudinal data analysis where the random intercept model has been applied to model the unobservable subject specific effects. However, it is quite likely that different independent sources of latent factors could be involved in $\mathbf{X}_u$. When $q > 1$, it is hard to assess if the regression coefficients from the two stages are proportional. In practice the number of latent variables is usually unknown. When likelihood methods are considered for estimation, some authors considered using an information criterion to determine the number of factors (Bai and Li (2014)). There are some limited discussions in econometrics and further study is still needed. On the other hand, when $q > 1$, the condition (C3) may still hold if $\mathbf{X}_u$ follows a multivariate normal distribution. The normality might be plausible for many biological characteristics symmetrically distributed across the population.

All simulations in Li et al. (2015) and Chan (2015) were carried out under the setting $q = 1$ and multivariate normality. Therefore the consistency of the 2SPS and 2SRI is witnessed. In this section we consider a few more general settings. In the following, we consider $q = 2$ with $\mathbf{X}_u = (X_{u1}, X_{u2})^T$. We assume X_e is generated by

$$X_e = 1 + 0.5X_I + X_o + .75X_{u1} + .5X_{u2} + .2\epsilon,$$

where X_I, X_o, X_{u1} and ϵ are all standard normal random variates. Three types of distributions for X_{u2} are considered: (A) uniform distribution over (1, 2); (B) log standard normal distribution; (C) Bernoulli distribution with success probability 0.5. The exact failure time is generated according to the following additive hazards model

$$h(t) = 9.5 + .5X_e + .5X_o + .75X_{u1} + 5.5X_{u2}.$$

The censoring time is generated from an exponential distribution with a hazard rate 2.5.

The results from 2SPS and 2SRI are summarized in Table 1. In the first case the performance of 2SPS appears to be slightly better than 2SRI. The 2SRI works well when the distribution of X_{u2} is uniform and only a slight departure from the normal. However, in the second case where the unobserved confounder follows a positively skewed distribution, the estimation bias based on 2SRI is much larger than that based on 2SPS. In the third case, we examine the performance when X_{u2} is discrete. It is noted that 2SRI yields higher bias than 2SPS with $n \leq 800$. When the sample size is extremely large, the binomial distribution converges to the normal distribution. The two methods are thus quite agreeable at $n = 1200$. In all cases, 2SRI seem to produce larger estimation variances than 2SPS.

Table 1: Performance of estimation results for $\hat{\beta}_e$ in 1000 simulations. Bias refers to the average biases over 1000 simulations. Var refers to the empirical variances of the estimates. Cov refers to the coverage of 95% confidence intervals.

		2SPS			2SRI		
Case	Sample size	Bias	Var	Cov (%)	Bias	Var	Cov (%)
A	200	-.136	8.842	93	-.1462	9.006	92
A	400	.133	4.191	93	.206	4.221	92
A	800	.163	2.003	91	.170	2.012	91
A	1200	.0244	.996	92	.0253	0.998	90
B	200	.803	10.14	92	10.15	50.53	29
B	400	.449	9.43	86	13.71	29.53	8
B	800	1.48	188	89	3.78	373	12
B	1200	-.562	71.5	90	8.003	138.8	14
C	200	-.176	3.92	96	-.217	3.97	95
C	400	-.0011	1.876	94	.0085	1.89	91
C	800	-.1478	.990	96	-.1585	.995	94
C	1200	.1065	.609	91	.0983	.610	91

5 Censoring assumption

We agree that the censoring assumption in Li et al. (2015) is stronger than that in Chan (2015). Many useful statistical methods in survival analysis depend critically on the underlying censoring assumption. It is strongly encouraged that practitioners examine how realistic to make the conditional independence assumption. Specifically, if X_e is categorical variable, one may adopt some goodness-of-fit test procedures to examine whether censoring time is independent of X_e. When the independence is acceptable, then application of structural equation modeling may be justified.

Acknowledgement

We acknowledge grants support from AcRF R-155-000-152-112, NMRC NMRC/CBRG/0014/2012 CBRG11nov060, and the Early Career Scheme from Hong Kong Research Grants Council [grant number PolyU 253023/16P]..

Appendix

Proof of Theorem 3.1 Note that

$$\begin{aligned} S(t; X_e, \mathbf{X}_o, v, \epsilon, \epsilon') &= S_0(t)\exp\{-t(\beta_e X_e + \boldsymbol{\beta}_o^T \mathbf{X}_o + \eta\} \\ &= S_0(t)\exp\{-t(\beta_e X_e + \boldsymbol{\beta}_o^T \mathbf{X}_o + \gamma_v v + \epsilon'\}. \end{aligned} \quad (5)$$

We have:

$$\begin{aligned} S(t; X_e, \mathbf{X}_o, v) &= \int_0^\infty S_0(t)\exp\{-t(\beta_e X_e + \boldsymbol{\beta}_o^T \mathbf{X}_o + \gamma_v v)\}\exp\{-\epsilon' t\} dF(\epsilon') \\ &= S_0^\dagger(t)\exp\{-t(\beta_e X_e + \boldsymbol{\beta}_o^T \mathbf{X}_o + \gamma_v v)\}, \end{aligned}$$

where $S_0^\dagger(t) = S_0(t)\int_0^\infty \exp\{-\epsilon' t\} dF(\epsilon')$ is a modified baseline survival function that does not involve $X_e, \mathbf{X}_o$ and v. Let $h_0^\dagger$ and $H_0^\dagger$ be the hazard and cumulative hazard function corresponding to $S_0^\dagger$ given above. For every i and t, the counting process $N_i(t)$ then admits the following unique decomposition:

$$N_i(t) = M_i(t) + \int_0^\infty A_i(s) dH^\dagger(s; X_{ei}, \mathbf{X}_{oi}),$$

where $H^\dagger(s; X_{ei}, \mathbf{X}_{oi}, v_i) = \int_0^\dagger h_0^\dagger(s) ds + t(\beta_{Te} X_{ei} + \boldsymbol{\beta}_{To}^T \mathbf{X}_o + \gamma_{Tv} v_i)$. Together with (4) we have

$$\begin{aligned} & U(\boldsymbol{\beta}) \\ = \ & \sum_{i=1}^n \int_0^\infty [\mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))] dM_i(t) + \\ & \sum_{i=1}^n \int_0^\infty [\mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))] A_i(t) dt \times [\beta_{Te} X_{ei} + \boldsymbol{\beta}_{To}^T \mathbf{X}_o + \gamma_{Tv} v_i - \boldsymbol{\beta}^T \mathbf{X}_{EOV} \mathbf{X}_{IOEi}]. \end{aligned}$$

Using the strong law of large numbers for martingale and the dominated convergence theorem we can show that n^{-1} of the first term in the right hand side of the above equality converges to zero almost surely. The second term is equal to

$$\sum_{i=1}^{n}\int_0^{\infty}[\mathbf{X}_{EOV}(\mathbf{X}_{IOEi}-\bar{\mathbf{X}}_{IOE}(t))]A_i(t)dt \times [(\beta_{Te}-\beta_e)X_{ei}+(\boldsymbol{\beta}_{To}-\boldsymbol{\beta}_o)^T\mathbf{X}_o+\gamma_{Tv}v_i-\gamma_v\hat{v}_i].$$

Under the assumed conditions and the law of large number, n^{-1} of the above equation converges almost surely to zero if and only if $\boldsymbol{\beta}=\boldsymbol{\beta}_T$. The strong convergence of $\hat{\boldsymbol{\beta}}$ is then obtained by noticing that it is the solution of an asymptotic unbiased estimating equation.

Proof of Theorem 3.2 From Taylor expansion we have:

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_T) &= \sqrt{n}\left\{\sum_{i=1}^{n}\int_0^{\infty}[\mathbf{X}_{EOV}(\mathbf{X}_{IOEi}-\bar{\mathbf{X}}_{IOE}(t))]^{\otimes 2}A_i(t)dt\right\}^{-1}U(\boldsymbol{\beta}_T)+o_p(1)\\ &\stackrel{d}{\to} n^{-1/2}\Omega^{-1}U(\boldsymbol{\beta}_T).\end{aligned}$$

Hence it suffices to show that $n^{-1/2}U(\boldsymbol{\beta}_T)$ is asymptotically normal with mean 0_{p+2} and covariance matrix $\Psi+\Sigma$.

Note that

$$\begin{aligned}n^{-1/2}U(\boldsymbol{\beta}_T) &= n^{-1/2}\sum_{i=1}^{n}\int_0^{\infty}[\mathbf{X}_{EOV}(\mathbf{X}_{IOEi}-\bar{\mathbf{X}}_{IOE}(t))]dM_i(t)+\\ &\quad n^{-1/2}\sum_{i=1}^{n}\int_0^{\infty}[\mathbf{X}_{EOV}(\mathbf{X}_{IOEi}-\bar{\mathbf{X}}_{IOE}(t))]A_i(t)dt\times[\gamma_{Tv}(v_i-\hat{v}_i)].\end{aligned}$$

Denote the two terms on the right hand side of the above equation as $\mathcal{I}_1$ and $\mathcal{I}_2$. Clearly, $\mathcal{I}_1$ is a martingale integral and is asymptotically normally distributed with zero mean and covariance matrix Ψ. For $\mathcal{I}_2$, note that for $i=1,\ldots,n$, by Conditions

(C5) and (C8) we have

$$
\begin{aligned}
& \lim_{n\to\infty} \Big\{ [\gamma_{Tv}(v_i - \hat{v}_i)]^2 \sum_{i=1}^{n} \int_0^\infty [\mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))] A_i(t) dt \\
& \times \mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))]^T A_i(t) dt \Big\}^{-1} \int_0^\infty [\mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))] A_i(t) dt \\
& \times [\gamma_{Tv}(v_i - \hat{v}_i)] \\
= \quad & \lim_{n\to\infty} \Sigma^{-1} n^{-1} \int_0^\infty [\mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))] A_i(t) dt \times [\gamma_{Tv}(v_i - \hat{v}_i)] = 0.
\end{aligned}
$$

By the Lindberg-Feller central limit theorem we know that $\mathcal{I}_2$ is asymptotically normally distributed with zero mean and covariance matrix Σ. Next we finish the proof by showing that the covariance between $\mathcal{I}_1$ and $\mathcal{I}_2$ converges to zero in probability. Note that

$$
\begin{aligned}
Cov(\mathcal{I}_1, \mathcal{I}_2) \quad = \quad & n^{-1} \sum_{i=1}^{n} Cov\Big(\int_0^\infty [\mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))] dM_i(t), \\
& \int_0^\infty [\mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))] A_i(t) dt \times [\gamma_{Tv}(v_i - \hat{v}_i)] \Big) \\
& + n^{-1} \sum_{1\le i \ne j \le n} Cov\Big(\int_0^\infty [\mathbf{X}_{EOV}(\mathbf{X}_{IOEi} - \bar{\mathbf{X}}_{IOE}(t))] dM_i(t), \\
& \int_0^\infty [\mathbf{X}_{EOV}(\mathbf{X}_{IOEj} - \bar{\mathbf{X}}_{IOE}(t))] A_j(t) dt \times [\gamma_{Tv}(v_j - \hat{v}_j)] \Big).
\end{aligned}
$$

By noticing that $\gamma_{Tv}(v_j - \hat{v}_j)$ converges to zero in probability for every $i = 1, \ldots, n$, we immediately have that $Cov(\mathcal{I}_1, \mathcal{I}_2) \xrightarrow{p} 0$.

Proof of Theorem 3.3 The 2SPS estimator of β can then be written as

$$\tilde{\beta} = \left\{ \frac{1}{n} \sum_{i=1}^{n} \int_0^\infty [\mathbf{X}_{si} - \bar{\mathbf{X}}_s(u)]^{\otimes 2} A_i(u) du \right\}^{-1} \times \left\{ \frac{1}{n} \sum_{i=1}^{n} \int_0^\infty [\mathbf{X}_{si} - \bar{\mathbf{X}}_s(u)] dN_i(u) \right\},$$

After some calculation it can be shown that the 2SRI estimator of β can be written as

$$\hat{\beta} = \left\{ \frac{1}{n} \sum_{i=1}^{n} \int_0^\infty [\mathbf{X}_{svi} - \bar{\mathbf{X}}_{sv}(u)]^{\otimes 2} A_i(u) du \right\}^{-1} \times \left\{ \frac{1}{n} \sum_{i=1}^{n} \int_0^\infty [\mathbf{X}_{svi} - \bar{\mathbf{X}}_{sv}(u)] dN_i(u) \right\},$$

Note that $\tilde{X}_{ei}$ is independent of v_i, and $\hat{X}_{ei}$ and $\hat{v}_i$ are root-n consistent. We have

$$\begin{aligned} \epsilon_{e,n} &= \frac{1}{n} \sum_{i=1}^{n} \int_0^\infty \left(\tilde{X}_{ei} - \frac{\sum_{j=1}^n A_j(u) \tilde{X}_{ej}}{\sum_{j=1}^n A_j(u)} \right) \left(v_i - \frac{\sum_{j=1}^n A_j(u) v_j}{\sum_{j=1}^n A_j(u)} \right) A_i(u) du + O_p(n^{-1/2}) \\ &= O_p(n^{-1/2}). \end{aligned}$$

Similarly, it can be shown that $\epsilon_{oj,n} = O_p(n^{-1/2})$ for $j = 1, \ldots, p$. Using the matrix inversion formula in block form we immediately have:

$$\begin{aligned} \hat{\beta}_e &= \tilde{\beta}_e + \alpha [B^{-1} C (B^{-1} C)^T]_{1,\cdot} (Y_{N,1}, \ldots, Y_{N,p+1})^T - \alpha (B^{-1} C)_1 Y_{N,p+2} \\ &= \tilde{\beta}_e + O_p(n^{-1}) + O_p(n^{-1/2}). \end{aligned}$$

Consequently, we have

$$Var(\hat{\beta}_e) = Var(\tilde{\beta}_e) + Var(\alpha (B^{-1} C)_1 Y_{N,p+2}) - 2Cov(\tilde{\beta}_e, \alpha (B^{-1} C)_1 Y_{N,p+2}) + O(n^{-3/2}).$$

References

Bai, J. and Li, K. (2014). Theory and methods of panel data models with interactive effects. *The Annals of Statistics*, 42:142–170.

Bijwaard, G. E. (2008). Instrumental variable estimation for duration data. Tinbergen Institute Discussion Paper.

Chan, K. C. G. (2015). Instrumental variable additive hazards models with exposure-dependent censoring. *Biometrics*.

Chen, S., Hsiao, C., and Wang, L. (2011). Measurement errors and censored structural latent variables models. *Econometric Theory*, 28:696–710.

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

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81:61–71.

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.

Muthen, B. and Masyn, L. (2005). Discrete-time survival mixture analysis. *Journal of Educational and Behavioral Statistics*, 30:27–58.

Posner, S. F. and Baker, L. (2000). Evaluation and extensions of a structural equation modeling approach to the analysis of survival data. *Behavior Genetics*, 30:41–50.

Robins, J. and Tsiatis, A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communication in statistics – theory and methods*, 20:2609–631.

Tang, M. L. and Lee, S. Y. (1998). Analysis of structural equation models with censored or truncated data via EM algorithm. *Computational Statistics and Data Analysis*, 27:33–46.

Terza, J. V., Basu, A., and Rathouz, P. J. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economic*, 27:531–543.