Estimating the Causal Effect of Treatment in Observational Studies with Survival Time Endpoints and Unmeasured Confounding

Jaeun Choi* A. James O'Malley†

Abstract

Estimation of the effect of a treatment in the presence of unmeasured confounding is a common objective in observational studies. The Two Stage Least Squares (2SLS) Instrumental Variables (IV) procedure is frequently used but is not applicable to time-to-event data if some observations are censored. We develop a simultaneous equations model (SEM) to account for unmeasured confounding of the effect of treatment on survival time subject to censoring. The identification of the treatment effect is assisted by IVs and the assumed bivariate distribution underlying the data generating process. The methodology is illustrated on data from an observational study of time to death following endovascular or open repair of ruptured abdominal aortic aneurysm. As the IV and the distributional assumptions cannot be jointly assessed from the observed data, we evaluate the sensitivity of the results to these assumptions.

Key Words: Comparative effectiveness research, Instrumental variable, Observational study, Simultaneous equations model, Survival analysis

1. Introduction

Instrumental Variable (IV) procedures such as Two Stage Least Squares (2SLS) are frequently used to account for unmeasured confounding, whose presence is a common concern in observational studies, with regard to estimation of the causal effect of treatment. However, IV procedures are not applicable to time-to-event data if some observations are censored. Thus, there is a clear need for an IV procedure that can be used with censored survival data as the exclusion of censored observations is subject to selection bias whenever there is differential loss to follow-up (Hernan, Hernandez-Diax, and Robins 2004). On the other hand, survival time models that ignore unmeasured confounding will yield biased treatment effects.

Although there recently has been some application of IV methods to survival analysis, they have not accounted for censoring, have employed methods that were not fully validated, or have provided inexplicit interpretation of treatment effect.

Specifically, Terza, Basu, and Rathouz (2008) showed that, for the nonlinear situations where 2SLS is less justified, Two Stage Residual Inclusion (2SRI) estimation is consistent across Weibull models of complete data and other nonlinear models. However, 2SRI does not account for censoring. In other work, O'Malley *et al.* (2011a) used a bivariate probit model (Heckman 1978; Goldman *et al.* 2001; Zeng *et al.* 2006; Bhattacharya, Goldman, and McCaffrey 2006) to jointly model both binary treatment and outcome by a simultaneous equation system; the equation for treatment is linked to an equation for the outcome through both treatment and a selection parameter that reflects the sign and magnitude of the correlation between the error terms of the two equations. Further, O'Malley, Frank, and Normand (2011b) modeled the causal effect of a treatment on a continuous outcome using a simultaneous equations model with IVs, but again censoring was not considered.

In the context of survival data subject to censoring, some studies applied the IV procedures but just illustrated their approaches through examples not validating whether they worked in the setting with censored observations: Abbring and van den Berg (2005) examined the empirical analysis of treatment effects on duration outcome from data including

*Department of Health Care Policy, Harvard Medical School, Boston, MA 02115

†The Dartmouth Institute of Health Policy & Clinical Practice, Geisel School of Medicine, Dartmouth College, Lebanon, NH 03766

instrumental variation in social experiments and considered a binary IV which is randomization to a binary treatment; Bosco *et al.* (2010) considered an IV-like estimation, similar to Brookhart and Schneeweiss (2007)'s preference-based IV method, by using the logistic regression in the first step and Cox proportional hazards regression in the second step; Recently, Gore *et al.* (2010) and Palmer (2013) used the control function approach of which 2SRI is a special case – Gore *et al.* (2010) applied 2SRI by considering Weibull survival models in the second stage, and Palmer (2013) employed a nonlinear IV approach in a proportional hazards model with endogeneity whose conditional distribution was assumed to be a normal distribution. However, these methods were not completely justified although there were some theories they proved.

As alternatives to the direct application of IVs in survival analysis, Blundell and Powell (2007) and Chernozhukov, Fernandez-Val, and Kowalski (2011) adopted IVs to account for unmeasured confounding in duration models, but their methods were based on quantile estimation of censored regression and thus treatment effect is the regression coefficient of treatment on conditional quantiles of outcome variable. Bijwaard (2009) proposed the Instrumental Variable Linear Rank Estimation (IVLRE) method for duration models that adjusts for the possible endogeneity of the intervention. The method used the inverse of the rank test for the significance of a covariate on the hazard and considered a Generalized Accelerated Failure Time (GAFT) model, but the causal effect in the GAFT model is defined in terms of shifting the quantiles of the outcome distribution. Furthermore, the IVLR method assumes that censoring time is (potentially) known in advance even for uncensored observations. Yu *et al.* (2012) evaluated a prior event rate ratio (PERR) adjustment method, and an alternative (PERR-ALT), in a Cox proportional hazards model. However, PERR and PERR-ALT rely on the availability of survival data from a prior time-period in order to account for baseline differences between the study groups in the absence of the intervention treatment and do not involve formal use of IVs. In these alternative approaches, the interpretation of treatment effect is not clearly understandable.

The lack of an established IV methodology for survival data motivates development of a new method to account for unmeasured confounding of the causal effect of treatment on survival time subject to censoring. For this purpose, we propose a structural equations model (SEM) by linking the survival time and treatment selection equations. We consider a log-normal survival model and assume the log-survival time and propensity to be "treated" have an underlying bivariate normal distribution; the method can thus be viewed as an extension of the bivariate probit model to survival time data. From the perspective of missing data, the method is a missing-not-at-random (MNAR) procedure as missingness depends on unobservables. The identification of the treatment effect is assisted by IVs that are related to treatment but conditional on treatment do not directly affect the outcome. Bayesian Markov Chain Monte Carlo (MCMC) methods are used for estimation. We suggest the imputation of censored survival times for computational efficiency in estimation and use a novel procedure to update estimated censored survival times over the MCMC iterations.

The methodology is illustrated on an example assessing endovascular (endo) repair vs open surgical (open) repair for patients with ruptured abdominal aortic aneurysm (rAAA), and compared to existing methods using non-censored samples or ignoring censored samples. Finite sample properties of the proposed method are investigated through simulation studies under various settings including strong and weak IVs and different censoring rates. In addition to when the model holds, we also conduct simulations to examine robustness under model misspecification.

The outline of this paper is as follows. We present our proposed approach and describe the estimation procedure in Section 2, apply the method to rAAA data in Section 3, provide numerical results from simulation studies in Section 4, and conclude in Section 5.

2. METHODS

2.1 Model Formulation and Notation

We denote the survival time, the right-censoring time, a treatment, a vector of exogenous covariates, and an unmeasured confounding variable for the i-th of n subjects by Y_i, C_i, W_i, $\boldsymbol{X}_i$, and U_i, respectively. While (Y_i, C_i) may be considered a bivariate random variable, we only observe $Y_i^{obs} = \min(Y_i, C_i)$, the followup time, and $\Delta_i = \mathrm{I}(Y_i \leq C_i)$, the status of Y_i^{obs}. We assume Y_i and C_i are independent conditional on $\boldsymbol{X}_i$ and U_i, and we further assume treatment is binary (e.g. endo versus open surgery) and Y_i^{obs} (i.e., Y_i and C_i) to have undergone a log transformation.

To present our assumed causal model we use potential outcomes notation. Let $Y_i(W)$ denote the potential survival time for subject i under treatment $W \in \{0, 1\}$. The survival time that would be observed in the absence of censoring satisfies the consistency equation, $Y_i = WY_i(1) + (1 - W)Y_i(0)$. The causal model for the survival times has the form

$$\mathrm{E}[Y_i(W)|\boldsymbol{X}_i, U_i] = \psi W + \boldsymbol{\beta}^T \boldsymbol{X}_i + \beta_u U_i. \tag{1}$$

The parameter, ψ, denoting the effect of treatment on survival time is the target of inference. A defining feature of the model assumed here is that W_i, the observed treatment for subject i, and U_i *are not* assumed to be independent.

Next we formalize the dependence of W_i on other variables, often referred to as selection variables, that influence the treatment an individual is assigned. We now introduce a vector of variables, denoted $\boldsymbol{Z}_i$, that are known as "instruments" as they manipulate treatment without fully controlling it and affect the outcome only indirectly through their manipulation of the treatment (Imbens and Angrist 1994). The potential treatment an individual is assigned can be represented by its own causal model. We specifically assume $W_i(\boldsymbol{Z}) = I(W_i^*(\boldsymbol{Z}) > 0)$, where

$$E[W_i^*(\boldsymbol{Z})] = \boldsymbol{\lambda}^T \boldsymbol{Z} + \boldsymbol{\theta}^T \boldsymbol{X}_i + \theta_u U_i, \tag{2}$$

for $\boldsymbol{Z} \in \mathcal{R}(\boldsymbol{Z})$, the set of all possible values of $\boldsymbol{Z}$. The quantity $W_i^*(\boldsymbol{Z})$ represents the underlying propensity on an unrestricted continuous scale of individual i to be treated (assigned $W_i = 1$) when $\boldsymbol{Z}_i = \boldsymbol{Z}$. The IV $\boldsymbol{Z}_i$ is thus assumed to (i) be associated with W_i conditional on $\boldsymbol{X}_i$ and U_i; (ii) have no effect on Y_i conditional on W_i, $\boldsymbol{X}_i$, and U_i; and (iii) be unrelated to U_i conditional on $\boldsymbol{X}_i$.

Together, the causal models in Equations 1 and 2 define the data generating mechanism under which (i) individuals are assigned (or select) a treatment and (ii) the observed and unobserved selection factors together with treatment affect outcomes. In order to estimate the parameters of these equations, and in particular ψ, in the presence of censoring we embed these causal equations in the following SEM:

$$Y_i = \psi W_i + \boldsymbol{\beta}^T \boldsymbol{X}_i + \epsilon_{1,i} \quad \text{and} \quad W_i^* = \boldsymbol{\lambda}^T \boldsymbol{Z}_i + \boldsymbol{\theta}^T \boldsymbol{X}_i + \epsilon_{2,i}, \tag{3}$$

where $W_i = I(W_i^* > 0)$, $\boldsymbol{\epsilon}_i = (\epsilon_{1,i}, \epsilon_{2,i}) = (\beta_u U_i + \delta_{1,i}, \theta_u U_i + \delta_{2,i})$, and $(\delta_{1,i}, \delta_{2,i})$ is a bivariate vector of independent random variables (independent of all other variables). Because $\boldsymbol{\epsilon}_i$ depends on U_i it follows that $\mathrm{Cov}(W_i, \epsilon_{1,i}) \neq 0$, which violates the conditions under which least squares or other methods for estimating linear regression models are valid.

To complete the specification of the model we assume

$$\boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{1,i} \\ \epsilon_{2,i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1 \\ \rho\sigma_1 & 1 \end{pmatrix} \right), \tag{4}$$

where σ_1^2 is interpreted as the residual variance of the survival time distribution. Because we only observe the binary treatment W_i, not the underlying propensity to be treated W_i^*, the restriction $\mathrm{var}(\epsilon_{2,i}) = 1$ is imposed in order to identify model parameters. The quantity ρ represents the net effect of unmeasured confounding by characterizing the extent to which

unobserved factors (U_i) affecting W_i^* are correlated with those affecting Y_i. A positive correlation ($\rho > 0$) indicates (in our case) endo-favorable selection because unobserved factors that make subjects more likely to receive endo increase survival time.

This model formulation is in similar fashion to the models by Chib and Hamilton (2000, 2002) assuming a multivariate normal distribution for continuous underlying treatment and potential outcomes whose observations are binary treatment and continuous outcomes.

The primary role of the log-normal assumption for survival time and other distributional restrictions in the bivariate normal distribution is to account for censoring and unmeasured confounding simultaneously in identification of the causal effect of treatment on survival time. Note that accounting for unmeasured confounding is also supported by the IVs included in the SEM.

2.2 Likelihood

Bayesian methods are used for estimation, so we must specify both the likelihood function for the observed data (this subsection) and a prior distribution for the unknown parameters (the next subsection).

For each subject we observe Y_i^{obs}, Δ_i and W_i conditional on $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$. Therefore, the likelihood function of the model parameters $(\psi, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\theta}, \sigma_1^2, \rho)$ given these observations is the product over $i = 1, \ldots, n$ of terms of the form:

$$
\begin{aligned}
& f\big(Y_i^{obs}, \Delta_i, W_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\big) \\
&= \left[\phi\big(Y_i^{obs}; \mu_{y,i}, \sigma_1^2 | W_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\big)\Phi\big(\mu_{w|y,i} | \boldsymbol{X}_i, \boldsymbol{Z}_i\big)^{W_i}\big(1 - \Phi\big(\mu_{w|y,i} | \boldsymbol{X}_i, \boldsymbol{Z}_i\big)\big)^{1-W_i}\right]^{\Delta_i} \\
&\quad \times \left[S\big(Y_i^{obs} | W_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\big) f\big(W_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\big)\right]^{1-\Delta_i}, \qquad (5)
\end{aligned}
$$

which are obtained by integrating the density of models (3) and (4) over W_i^* (i.e. integrating over $\epsilon_{2,i}$). The components of (5) are given by:

$$
\mu_{y,i} = \psi W_i + \boldsymbol{\beta}^T \boldsymbol{X}_i, \qquad \mu_{w|y,i} = \frac{\boldsymbol{\lambda}^T \boldsymbol{Z}_i + \boldsymbol{\theta}^T \boldsymbol{X}_i + \rho(Y_i^{obs} - \mu_{y,i})/\sigma_1}{\sqrt{1-\rho^2}}, \qquad (6)
$$

$$
S\big(Y_i^{obs} | W_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\big) = P\big(Y_i \geq Y_i^{obs} | W_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\big) = \int_{Y_i^{obs}}^{\infty} \phi\big(Y_i; \mu_{y,i}, \sigma_1^2 | W_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\big) dY_i, \text{ and}
$$

$$
f\big(W_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\big) = \Phi\big(\mu_{w^*,i} | \boldsymbol{X}_i, \boldsymbol{Z}_i\big)^{W_i}\big(1 - \Phi\big(\mu_{w^*,i} | \boldsymbol{X}_i, \boldsymbol{Z}_i\big)\big)^{1-W_i} \text{ with } \mu_{w^*,i} = \boldsymbol{\lambda}^T \boldsymbol{Z}_i + \boldsymbol{\theta}^T \boldsymbol{X}_i,
$$

where $\phi(v; \mu, \sigma^2)$ and $\Phi(v)$ denote the probability density function of $N(\mu, \sigma^2)$ and the cumulative density function of $N(0, 1)$, respectively, evaluated at v, for $\rho \neq 1$.

2.2.1 Parameter expansion to censored survival times

The estimation procedure based on the likelihood (5) is complicated due to the joint problem of censoring and unmeasured confounding: in survival analysis, a censored subject (i.e. $Y_i^{obs} = C_i$ and $\Delta_i = 0$) contributes to the likelihood through survival probability at the observed censoring time; with unmeasured confounding, the survival function $S\big(Y_i^{obs} | W_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\big)$ in (5) is obtained by integrating its conditional survival function over unmeasured confounders; when the function does not have a closed form, this integration becomes too complex and computationally intensive like our SEM. Therefore, instead of using survival probability, we propose to impute censored survival times truncated by the corresponding observed censoring times, described in Section 2.4.1, which simplifies the computation which would have been done by the original complicated estimation procedure.

For this imputation, we expand parameters to be estimated by treating the censored survival time as a range restricted unknown parameter denoted Y_i^*. Then, the conditional complete data likelihood function for $(\psi, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\theta}, \sigma_1^2, \rho, Y_i^* | Y_i = Y_i^* \geq C_i)$ is given by:

$$\left[\phi\left(Y_i;\mu_{y,i},\sigma_1^2|W_i,\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\Phi\left(\mu_{w|y,i}|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)^{W_i}\left(1-\Phi\left(\mu_{w|y,i}|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^{1-W_i}\right]^{\Delta_i}$$
$$\times\left[\phi\left(Y_i^*;\mu_{Y^*,i},\sigma_1^2|W_i,\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\Phi\left(\mu_{w|y^*,i}|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)^{W_i}\left(1-\Phi\left(\mu_{w|y^*,i}|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^{1-W_i}\right]^{1-\Delta_i}$$
$$=\phi\left(Y_i^a;\mu_{Y^a,i},\sigma_1^2|W_i,\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\Phi\left(\mu_{w|y^a,i}|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)^{W_i}\left(1-\Phi\left(\mu_{w|y^a,i}|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^{1-W_i}, \quad (7)$$

where $Y_i^a = Y_i^{\Delta_i}(Y_i^*)^{(1-\Delta_i)}$, $\mu_{Y^*,i}$ and $\mu_{w|y^*,i}$ have Y_i^* in place of Y_i in (6), and $\mu_{Y^a,i}$ and $\mu_{w|y^a,i}$ have Y_i^a in place of Y_i in (6).

2.3 Prior and Posterior Distribution

The posterior distribution of the parameters $\Theta = (\psi, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\theta}, \sigma_1^2, \rho, \boldsymbol{Y}^*)$ in (7) is given by

$$P(\Theta|Data) = \frac{P(Data|\Theta)P(\Theta)}{\int P(Data|\Theta)P(\Theta)d\Theta} = \frac{P(Data|\Theta)P(\Theta)}{P(Data)} \propto P(Data|\Theta)P(\Theta). \quad (8)$$

To further reduce computational complexity, we re-parameterize $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$, and ρ in (6) to parameters for which closed-form posterior distributions are more accessible via:

$$\tilde{\boldsymbol{\lambda}} = \frac{\boldsymbol{\lambda}}{\sqrt{1-\rho^2}}, \quad \tilde{\boldsymbol{\theta}} = \frac{\boldsymbol{\theta}-\boldsymbol{\beta}\rho/\sigma_1}{\sqrt{1-\rho^2}}, \quad \text{and} \quad \tilde{\rho} = \frac{\rho}{\sigma_1\sqrt{1-\rho^2}}. \quad (9)$$

Then, we specify priors for the censored survival outcome (Y_i^*), three original parameters $(\psi, \boldsymbol{\beta}, \sigma_1^2)$, and three transformed parameters $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\theta}}, \tilde{\rho})$. We assume diffuse priors for Y_i^* and the model parameters other than ρ. Because we are interested in sensitivity to unmeasured confounding, prior distributions for ρ of varying precision are used. Non-informative flat priors are assumed for Y_i^* over $[C_i, \infty)$, ψ, $\boldsymbol{\beta}$, $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\boldsymbol{\theta}}$. The Inverse Gamma prior, which is a conjugate prior for the normal distribution, is assumed for σ_1^2. Specifically, $p(Y_i^*) \propto 1$, $p(\psi) \propto 1$, $p(\boldsymbol{\beta}) \propto 1$, $p(\tilde{\boldsymbol{\lambda}}) \propto 1$, $p(\tilde{\boldsymbol{\theta}}) \propto 1$, and $\sigma_1^2 \sim IG(\upsilon_1, \upsilon_2)$, where υ_1 and υ_2 are chosen such that the variance far exceeds the mean. We assume a Beta prior for $\rho^* = (\rho+1)/2$, (i.e. $\rho^* \sim Beta(\nu_1, \nu_2)$), which expands to a Beta-type prior for ρ over $(-1, 1)$. Thus, ρ has the prior density:

$$p(\rho;\nu_1,\nu_2) = \left(\frac{1}{2}\right)^{\nu_1+\nu_2-1}\frac{1}{B(\nu_1,\nu_2)}(1+\rho)^{\nu_1-1}(1-\rho)^{\nu_2-1},$$

where $B(\nu_1, \nu_2) = \Gamma(\nu_1)\Gamma(\nu_2)/\Gamma(\nu_1+\nu_2)$. Therefore, $\tilde{\rho}$ has the prior,

$$p(\tilde{\rho},\nu_1,\nu_2) = \left(\frac{1}{2}\right)^{\nu_1+\nu_2-1}\frac{1}{B(\nu_1,\nu_2)}\frac{\sigma_1}{\sqrt{1+\tilde{\rho}^2\sigma_1^2}}\left(1+\frac{\tilde{\rho}\sigma_1}{\sqrt{1+\tilde{\rho}^2\sigma_1^2}}\right)^{\nu_1-1}\left(1-\frac{\tilde{\rho}\sigma_1}{\sqrt{1+\tilde{\rho}^2\sigma_1^2}}\right)^{\nu_2-1}. \quad (10)$$

In the special case where $\nu_1 = \nu_2 = 1$, ρ and $\tilde{\rho}$ follow uniform$(-1, 1)$ and t with 2 degrees of freedom, respectively. The conditional posterior distributions needed for the computation described in Section 2.4 are given in Appendices. We refer to the model given in (3) and (4) and the above prior distributions as the Bayesian structural equations model (BSEM).

2.4 Bayesian Computation

We briefly outline the Markov Chain Monte Carlo (MCMC) procedure used to fit the BSEM. For the $n_0 \leq n$ subjects with right-censored survival times, let $\boldsymbol{Y}_0^*$ denote a $n_0 \times 1$ vector of censored survival outcomes with the j-th element, $Y_{0,j}^*$, $j = 1, \ldots, n_0$. (Recall that $Y_i^* \geq C_i$ denotes the parameter for the censored survival outcome for subject i if $\Delta_i = 0$.)

The MCMC procedure first selects initial values $\left(\boldsymbol{Y}_0^{*(0)} = (Y_{0,1}^{*(0)}, \ldots, Y_{0,n_0}^{*(0)}), \psi^{(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{\lambda}^{(0)}, \boldsymbol{\theta}^{(0)}, \sigma_1^{2(0)}, \rho^{(0)}\right)$. They are updated to obtain $Y_{0,1}^{*(1)}, \ldots, Y_{0,n_0}^{*(1)}, \boldsymbol{\beta}^{(1)}, \sigma_1^{2(1)}, (\boldsymbol{\lambda}^{(1)}, \boldsymbol{\theta}^{(1)}, \rho^{(1)})$, and $\psi^{(1)}$. For $(\boldsymbol{\lambda}^{(1)}, \boldsymbol{\theta}^{(1)}, \rho^{(1)})$, their transformed counterparts $(\tilde{\boldsymbol{\lambda}}^{(1)}, \tilde{\boldsymbol{\theta}}^{(1)}, \tilde{\rho}^{(1)})$ are jointly updated and then transformed back to update the original parameters. $\boldsymbol{\beta}^{(1)}$ and $\sigma_1^{2(1)}$ are directly generated from their corresponding conditional posteriors, a normal and an Inverse-Gamma distribution, respectively, using Gibbs sampling. Candidate values for $Y_{0,1}^{*(1)}, \ldots,$

$Y_{0,n_0}^{*(1)}$, $(\tilde{\boldsymbol{\lambda}}^{(1)}, \tilde{\boldsymbol{\theta}}^{(1)}, \tilde{\rho}^{(1)})$, and $\psi^{(1)}$ are generated from the corresponding conditional posteriors by the Metropolis-Hastings (M-H) algorithm and each current value is updated with the newly generated value if its candidate is accepted based on the M-H acceptance probability. Otherwise, the current value is retained. The procedure is then iterated.

We found that 40,000 samples following burn-in of 10,000 samples were sufficient by studying and evaluating trace plots of the MCMC procedure, the Gelman-Rubin diagnostic, and the Geweke criterion for monitoring convergence of the MCMC procedure. The estimates of the parameters are the posterior means of the $M(= 40,000)$ samples.

2.4.1 Imputing censored survival times

In Section 2.2.1 we represented the censored survival times as range restricted unknown parameters to avoid computing their marginal survival probabilities, $S\big(Y_i^{obs}|W_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\big)$, which are computationally demanding. In the course of our MCMC algorithm we use a novel procedure to update the parameter Y_i^*, representing the true value of Y_i. We now justify this procedure.

When the survival time for the i-th subject is observed ($Y_i^{obs} = Y_i$), the joint marginal density for (Y_i, W_i) evaluated at (Y_i^{obs}, W_i) given in (5) is

$$\begin{aligned} &f(Y_i^{obs}, W_i|\boldsymbol{X}_i, \boldsymbol{Z}_i) \\ &= \phi\big(Y_i; \mu_{y,i}, \sigma_1^2|W_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\big)\Phi\big(\mu_{w|y,i}|\boldsymbol{X}_i, \boldsymbol{Z}_i\big)^{W_i}\big(1 - \Phi\big(\mu_{w|y,i}|\boldsymbol{X}_i, \boldsymbol{Z}_i\big)\big)^{1-W_i}. \quad (11) \end{aligned}$$

When the i-th subject is censored ($Y_i^{obs} = C_i$), Y_i^*, the realization of the survival time Y_i that would have been observed in the absence of censoring, is left-truncated by C_i. Given (11), $Y_i^* \geq C_i$, and based on an improper uniform prior for Y_i^* we approximate the conditional posterior distribution for Y_i^* given in Appendix A and use the approximate distribution to propose candidate draws of Y_i^* in the MCMC algorithm used to fit the BSEM.

To update Y_i^* we use the truncated normal distribution, $f(Y_i^*|W_i, \boldsymbol{X}_i, \boldsymbol{Z}_i) \propto \phi\big(Y_i^*; \mu_{Y^*,i}, \sigma_1^2|W_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\big)I(Y_i^* \geq C_i)$, as a candidate generating density and apply the M-H algorithm. Since $\mu_{Y^*,i} = \psi W_i + \boldsymbol{\beta}^T \boldsymbol{X}_i$ does not depend on Y_i^*, at each iteration the candidates are drawn independently of the current value of Y_i^* (an independence chain, Tierney 1994), in contrast to a random walk chain. Under this strategy, the probability of accepting a new candidate value $Y_i^{*,new}$ is given by

$$\min\left\{1, \frac{\Phi\big(\mu_{w|y^*,new,i}|\boldsymbol{X}_i, \boldsymbol{Z}_i\big)^{W_i}\big(1 - \Phi\big(\mu_{w|y^*,new,i}|\boldsymbol{X}_i, \boldsymbol{Z}_i\big)\big)^{1-W_i}}{\Phi\big(\mu_{w|y^*,i}|\boldsymbol{X}_i, \boldsymbol{Z}_i\big)^{W_i}\big(1 - \Phi\big(\mu_{w|y^*,i}|\boldsymbol{X}_i, \boldsymbol{Z}_i\big)\big)^{1-W_i}}\right\}.$$

We found the acceptance rate to be very high (around 0.9), which implies the shape of the truncated normal distribution used to generate posterior samples is close to the shape of the true conditional posterior for Y_i^*.

To examine the predictability of Y_i^* determined by the fitted model, we performed simulations and drew a plot of the density of log(estimated incremental survival time beyond censoring time) by treatment over censored observations from a simulated data set (Figure 1). The vertical lines in Figure 1 indicate the average of log(true incremental survival time beyond censoring time) over the censored observations in the corresponding treatment group. Figure 1 shows that, within each treatment group, the estimated log(incremental survival times beyond censoring times) are centered around the average of corresponding true values suggesting the imputations are well calibrated in terms of bias.

3. APPLICATION TO VASCULAR SURGERY DATA

In this section, the BSEM is illustrated with a clinical example using vascular surgery data and compared to the results under the standard survival analysis ignoring confounding and

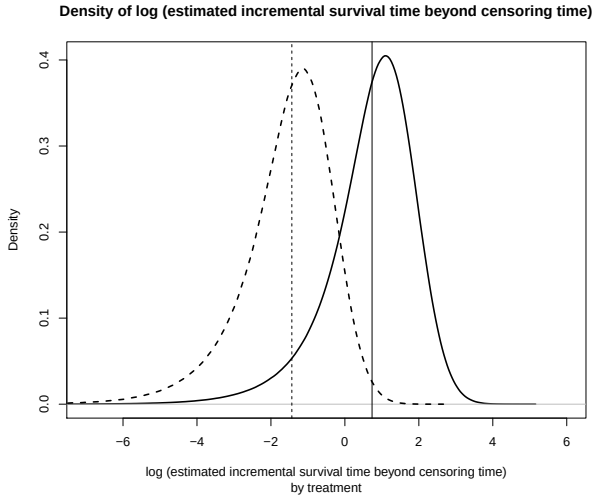


Figure 1: Density plot of log(estimated incremental survival time beyond censoring time). Vertical line indicates the average log(true incremental survival time beyond censoring time) for censored observations within treatment group, solid and dottd lines denote treatment and control treatment group, respectively, n = 5000, censoring rate = 30%, the number of MCMC posterior samples = 40000, ψ = 2, ρ = 0.2, and a vague prior for ρ was assumed.

the 2SLS procedure only applied to the non-censored subset of observations. As the standard survival analysis, we fit the log-normal accelerated failure time (AFT) model which has the same interpretation of treatment effect as the ones in the other methods.

Endovascular (endo) repair was introduced in 1999 as a less invasive alternative to elective open surgical (open) repair of AAA, which was traditionally performed to prevent ruptures. We evaluate its effectiveness on survival of patients with rAAA cases, one of the most fatal surgical emergencies, compared to that of open repair.

Medicare claims data were used to identify all open and endo repairs of rAAA that occurred during 2001–2008. To be eligible for analysis, patients were required to have at least 2 years of prior Medicare enrollment. This restriction ensured that comorbidities that might influence the choice of approach and outcomes of rAAA repair could be measured equitably for all patients. A total of 2,853 ruptured cases met the criteria for analysis, yielding 2,201 (77.15%) and 652 (22.85%) patients who underwent endo and open repair, respectively. The proportion of censored survival times is high (66.28%). Patients in endo group could have been followed up an average of 1578.53 days (SD=779.91) while those in open repair group could have been followed up an average of 1854.24 days (SD=886.56).

The survival time is evaluated as the number of days from procedure date to death. Treatment is coded 1 for endo and 0 for open repair. Based on a prior study (O'Malley *et al.* 2011a), we adjusted for the following observed confounders on which both treatment selection and survival may depend: gender, race, age at procedure, seven indicators of specific comorbidities (previous renal failure, congestive heart failure, chronic pulmonary disease, peripheral vascular disorders, vascular disease, neurovascular disease, and prior AAA diagnosis), two procedural factors (year of procedure and urgent case – defined as emergency department charges of $50 or more), and total number of AAA procedures over the prior 365 days at the hospital where the procedure was performed. There are suspected to be unmeasured confounders that influence choice of procedure.

The proportion of endo cases over the prior 365 days of each procedure at the hospital the patient attended is used as an IV. The rationale behind this choice is that the likelihood a rAAA patient receives endo is likely to increase with the proportion of endo cases performed at the hospital over the prior 365 days but, conditional on the patient's treatment

Table 1: Results of estimates of ψ, the effect of endovascular (endo) versus open surgical repair (open) for the treatment of ruptured abdominal aortic aneurysm (rAAA), in the Medicare population over 2001–2008.

		Treatment effect (ψ)		
Censored data	Procedure	Estimate	Std Err	Interval
Exclude	AFT model	.230	.060	(.112, .347)
	2SLS	.213	.135	(- .052, .477)
	BSEM	.283	.135	(.034, .577)
	BSEM w/o IV	1.168	.189	(.793, 1.529)
Include	AFT model	.076	.063	(- .048, .200)
	BSEM	- .112	.129	(- .353, .145)
	BSEM w/o IV	- .122	.238	(- .529, .333)

Note: The BSEM used a vague extended Beta prior with parameters=(1,1) for ρ (i.e. $\mathrm{E}[\rho]=0$ and $\mathrm{Var}[\rho]=.33$) over a range (-1,1) which is same as unif(-1,1).

and other covariates including the total number of AAA cases and the date of their procedure, it is unlikely that the proportion of endo cases (on other patients) will be correlated with the patient's survival time.

Since the bivariate normality assumed in the BSEM is not testable, we examined the univariate log-normality of survival time of the rAAA patients through the Q-Q plot of the standardized residuals by the the log-normal AFT model fitted to survival time as an exploratory approach to check the violation of the assumption. The Q-Q plot (not presented) showed rough indication of log-normality of univariate survival time and it did not look problematic to consider the log-normal survival model.

For the BSEM, we assume the prior mean of ρ to be 0 because (i) we are not knowledgeable about whether unmeasured confounders between treatment and outcomes are positively or negatively correlated in this rAAA data and using the prior center=0 allows both directions in estimation and (ii) ρ is not likely to be away from 0 since many observed confounders are already adjusted for in this analysis.

Table 1 compares the estimates of treatment effect (ψ) between estimation methods when censored observations are accounted for and excluded. Estimates are obtained using the BSEM with and without an IV, the log-normal AFT model (ignores confounding), and the 2SLS procedure (excludes censored cases). For comparison, the BSEM and the AFT model are also evaluated when censored observations are excluded. For the BSEM in this table, we assume a vague extended Beta prior with parameters=(1,1) for ρ with mean=0 and variance=0.33 over a range (-1,1) which is same as uniform (-1,1).

There are several informative findings in Table 1. First, the BSEM and the AFT model produce estimates of ψ that are smaller in magnitude when censoring is accounted for than with only the non-censored data. Thus, ψ is likely overestimated by removing censored observations. Second, ignoring confounding leads to estimates of smaller magnitude of ψ in the standard AFT model than when accounting for unmeasured confounding in the BSEM irrespective of whether censoring is accounted. However, the 2SLS yields the smallest estimated ψ among the methods for non-censored data. This may reflect a selection bias incurred from excluding the large amount of censored observations in the rAAA data. Third, the BSEM accounting for both censoring and unmeasured confounding yields negative point estimates of ψ indicating shorter survival under endo while all the other point estimates of ψ are positive, suggesting longer survival under endo. However, when accounting for censoring, the 95% credible intervals from the standard AFT model and the BSEM reassuringly include 0, implying endo neither significantly increased nor decreased

Table 2: Results of estimates of treatment effect (ψ) and correlation (ρ) by omitting/including 'urgent admission' among observed confounders for all rAAA data in the Medicare population over 2001–2008.

Confounders	Procedure	Parameter	Estimate	Std Err	Interval
Include all	AFT model	ψ	.076	.063	(-.048, .200)
observed confounders	BSEM	ψ	-.112	.129	(-.353, .145)
		ρ	.119	.071	(-.014, .259)
Omit an observed	AFT model	ψ	.095	.063	(-.029, .219)
confounder	BSEM	ψ	-.135	.144	(-.436, .133)
(urgent admission)		ρ	.145	.080	(-.009, .309)

Note: The BSEM used a vague extended Beta prior with parameters=(1,1) for ρ (i.e. $\mathrm{E}[\rho]=0$ and $\mathrm{Var}[\rho]=.33$) over a range (-1,1) which is same as unif(-1,1).

survival compared to open. Fourth, the BSEM without an IV appears to magnify the point estimate of ψ in the same direction of that by the BSEM with an IV and produces the wider interval estimates than the BSEM with an IV, indicating that including an IV in the BSEM helps the identification of ψ and stabilizes variance in the estimation of ψ.

As an ad hoc analysis to calibrate the impact of unmeasured confounders in the rAAA data, we created an unmeasured confounder by omitting an observed variable whose direction and magnitude of confounding can be assessed and compared the estimates to the results under the model including the omitted confounder. We chose to omit 'urgent admission', an indicator of whether a patient is urgently admitted, which appeared to negatively affect both treatment and outcome in the relatively larger magnitude than the other confounders (the estimates of coefficients of the observed confounders not presented in this paper). This analysis was conducted to all rAAA data including censored observations and we used a vague extended Beta prior for ρ as done for Table 1. The results from this analysis are provided in Table 2: (i) the magnitude of the estimate of ψ increases in both the AFT model and the BSEM when omitting 'urgent admission' from when adjusting for it, which implies the created unmeasured confounder caused ψ to be more overestimated by 0.019 positively in the AFT model and by 0.023 negatively in the BSEM; (ii) by the BSEM without 'urgent admission', ρ has the bigger point estimate and wider interval estimate which reflect the increased unmeasured confounding. Specifically, we may imply the increment=0.026 in the estimates of ρ which represents the net effect of the additional unmeasured confounder led to the increment=0.023 in the magnitude of the estimates of ψ.

We also assessed the sensitivity of the BSEM to different levels of precision of the prior for ρ, and the results when accounting for censoring are given in Table 3. The BSEM appears insensitive to the different priors for ρ, implying the vague prior has sufficient precision to be used on the rAAA data.

On the other hand, from the flipped sign of the point estimate of ψ between the AFT model and the BSEM when accounting for censoring (Table 2), one may question on the appropriateness of the BSEM to the rAAA example. Answering this question involves the evaluation of ρ. As mentioned earlier in this section, it is unlikely that unmeasured variables could exist such that ρ is far from 0 because we have a very detailed data set and AAA has been studied extensively. Furthermore, in Copas and Li (1997), it is suggested that ρ estimated to be near ± 1 indicates model lack-of-fit. From Table 3, reassuringly all the estimates of ρ under the BSEM are near 0 and their 95% credible intervals include 0. Hence, we may conclude the BSEM incorporating censored observations as well as unmeasured confounders yields believable and reasonable results.

In this application, the unmeasured confounders appear to have a big impact on the estimation of ψ even with the small ρ estimated to be around 0.1. Thus, in Section 4, we

Table 3: Results of the BSEM estimates of treatment effect (ψ) and correlation (ρ) for all rAAA data in the Medicare population over 2001–2008.

Parameter	Prior of ρ	Estimate	Std Err	Interval
ψ	Vague	-.112	.129	(-.353, .145)
	Medium	-.094	.135	(-.367, .175)
	Precise	-.067	.130	(-.327, .176)
ρ	Vague	.119	.071	(-.014, .259)
	Medium	.108	.075	(-.044, .254)
	Precise	.092	.071	(-.042, .234)

Note: The vague, medium, and precise priors for ρ are Beta-type distributions with parameters=(1,1), (5,5) and (10,10), respectively. Thus, $\mathrm{E}[\rho] = 0$ in each case and $\mathrm{Var}[\rho]$ = .33, .09, and .05, respectively, over a range (-1,1).

further investigate the performance of the proposed method under various scenarios whose simulations are to be based on the larger range of values on the parameter space of ρ (i.e. $0/ \pm 0.2/ \pm 0.5$) than the realistic values shown in Tables 2 and 3 in order to find more explicit impact on the estimation of ψ by the change of ρ.

4. SIMULATION STUDIES

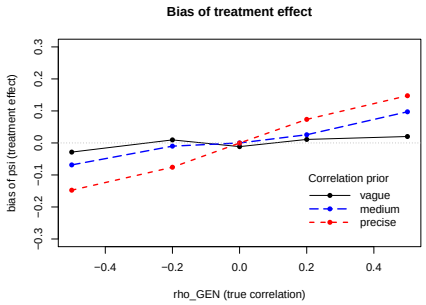
In this section, we conduct simulation studies to evaluate the sensitivity of $\hat{\psi} = E[\psi|\mathrm{Data}]$, the posterior mean estimator of ψ, to: (1) the prior of ρ, (2) strength of the IV, (3) different censoring rates, and (4) misspecification of the distribution of $\boldsymbol{\epsilon}_i$.

We assume X_i and Z_i are univariate and generated from uniform$(0, 1)$. Further, $\boldsymbol{\epsilon}_i = (\epsilon_{1,i}, \epsilon_{2,i})$ has a bivariate normal distribution except in those simulations evaluating the impact of a wrongly assumed error distribution. W_i^* is generated by the linear model in (3) (in Section 2.1) and W_i is 1 for $W_i^* > 0$ and 0 otherwise. The transformed outcome Y_i is generated by the linear model in (3); the exponential of Y_i can be thought of as the original positively-valued survival time under a log-normal model. The exponential of C_i is generated from uniform$(0, a)$ and log-transformed to C_i. The censoring proportion=30% except when evaluating the performance of the estimator at different censoring rates. Performance is evaluated by computing bias, mean squared error (MSE), and coverage probabilities over 100 simulated data sets, but results are presented only by bias plots instead of large arrays of tables (available from authors) since trends are obvious.

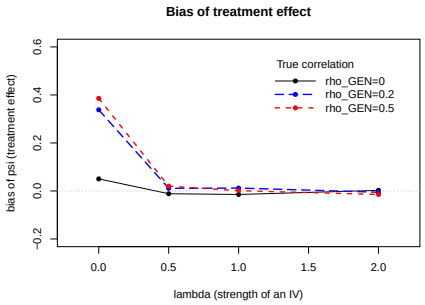
4.1 Finite Sample Properties and Sensitivity to ρ

The values of the survival-time parameters (Equation 1) for data generation were $\psi = -0.5$, $\beta_0 = 0.5$, and $\beta_1 = 0.5$ and the same for the treatment-selection parameters (Equation 2) were $\lambda = 0.5$, $\theta_0 = -0.5$ and $\theta_1 = 0.5$. The variance and correlation parameters in the bivariate normal error distribution (4) were set to $\sigma_1^2 = 0.5$ and $\rho_{\mathrm{GEN}} = 0/ \pm 0.2/ \pm 0.5$, where ρ_{GEN} distinguishes the value of ρ at which the estimation procedure is evaluated from the model (e.g., the prior distribution) used for the analysis.

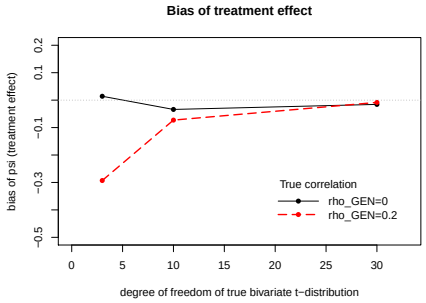
Beta-type distributions with parameters=(1,1), (5,5) and (10,10) form the vague (uniform or flat), medium, and precise priors of ρ, respectively, with prior mean=0 (no unmeasured confounding) and variance=0.33, 0.09, and 0.05. The choice of the prior mean=0 of ρ in estimation is reasonable because usually we are not knowledgeable about whether ρ is positive or negative in real application and using the prior center=0 allows both directions in estimation. However, for a particular case where researchers believe that there exists unmeasured confounding and are knowledgeable about the relationship of the unmeasured



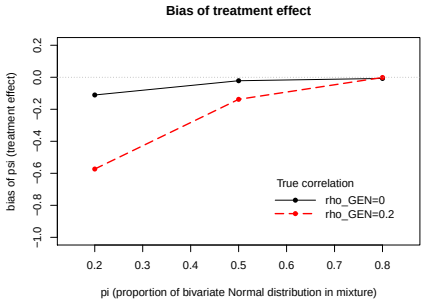
(a) Sensitivity to the prior of correlation (ρ) and true correlation (ρ_{GEN})



(b) Sensitivity to the strength of an IV



(c) Sensitivity to model misspecification with true bivariate t-dist'n



(d) Sensitivity to model misspecification with true mixture dist'n

Figure 2: Bias plots of treatment effect (ψ) in simulation studies

confounders between treatment and outcome, the prior of ρ should be centered according to the known information. Hence, we further conduct the simulations with the prior mean=0.2 of ρ for the situation assuming the prior mean $\neq$ 0 (existence of unmeasured confounding) under both true conditions (no unmeasured confounding / existence of unmeasured confounding). In this setting, the parameters of the Beta-type distributions were determined by the specified prior mean and variances of ρ to have vague (uniform or flat), medium, and precise priors. Two sample sizes n =1000 and 5000 are simulated and the result with n =5000 is presented in this paper.

Figure 2(a) shows the bias plot of ψ under the BSEM assuming no unmeasured confounding (the prior mean=0 of ρ) in estimation. From Figure 2(a), the bias of $\hat{\psi}$, the posterior mean estimator of ψ, increases as ρ_{GEN} moves away from the prior mean, $E[\rho]$=0, and the prior of ρ is more informative, and, conversely, bias is very close to 0 when $\rho_{GEN} = E[\rho]$ regardless of informativeness of prior. Also, (although not presented), MSE and coverage improved when $\rho_{GEN} = E[\rho]$ whereas they deteriorated $\rho_{GEN} \neq E[\rho]$, but informative priors for ρ had less influence as n increases from 1000 to 5000, showing bias and MSE decrease and the coverage probability converges upon the 95% nominal level.

The result from additional simulations by the BSEM in the presence of unmeasured confounding (the prior mean=0.2 of ρ) confirmed the result under no unmeasured confounding (the prior mean=0 of ρ) – bias close to 0 when $\rho_{GEN} = E[\rho]$ and more bias for the further ρ_{GEN} from the prior mean and for the more informative prior of ρ.

4.2 Sensitivity to the Strength of IVs

To illustrate the importance of a strong IV to the BSEM, we evaluate the operating characteristics of $\hat{\psi}$ at λ = 0, 0.5, 1, and 2 (the effect of the IV). A vague prior for ρ is assumed so that the IV alone is principally responsible for identifying ψ separate from ρ. The other

model specifications and simulation settings are as in Section 4.1.

Figure 2(b) reveals that the estimate of ψ is less biased when λ is bigger (stronger IV). It is clear that the benefit of a strong IV is relatively more evident for larger ρ_{GEN} due to the fact that the higher correlation between the survival time and the treatment selection equations increases the information available to estimate ψ provided the presence of unmeasured confounding is accounted.

4.3 Sensitivity to Censoring Rate

To assess sensitivity to different censoring proportions we set the censoring rate to be 0, 30%, and 60% while fixing the other simulation parameters to the same values as previously and assuming the uniform prior for ρ.

The bias plot of this simulation is not presented in Figure 2 because bias is close to 0 for all censoring rates. Although lower censoring rates led to better precision, the differences were very small among censoring rates. This reveals that the information in the censoring times makes an important contribution to estimation; even in the presence of substantial censoring precision is only modestly compromised.

4.4 Sensitivity to Model Misspecification

We finally examine the impact of departures from the bivariate normal assumptions. Data are generated assuming the true distribution of the random errors of the survival time and treatment selection equations to be (i) a bivariate t-distribution and (ii) a mixture of a bivariate normal distribution and a bivariate Gamma distribution. These allow sensitivity to outliers and skewness to be assessed, respectively.

4.4.1 Bivariate t-distribution

Under a bivariate t-distribution, the random errors in the survival and treatment selection equations have the PDF given by

$$\boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{1,i} \\ \epsilon_{2,i} \end{pmatrix} \sim t_{df}\left(\boldsymbol{\mu}_{\boldsymbol{\epsilon}}, \boldsymbol{\Sigma}\right), \quad \text{where} \quad \boldsymbol{\mu}_{\boldsymbol{\epsilon}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1 \\ \rho\sigma_1 & 1 \end{pmatrix}. \quad (12)$$

The smaller the degree-of-freedom (df) the heavier tails of the resulting PDF. We consider df=3, 10, and 30 and evaluate the operating characteristics of $\hat{\psi}$ when ρ_{GEN}=0 and 0.2 assuming the uniform prior for ρ. Other model specifications and simulation settings are as in Section 4.1.

As df increases and the true distribution approaches normality, we find the BSEM estimator of ψ is less biased (Figure 2(c)), more precise and has better coverage (not presented). Hence, the combination of unmeasured confounding (i.e. under $\rho_{GEN} = 0.2$) and assuming a survival time distribution whose tails are not sufficiently thick leads to substantial bias whereas the bias is small regardless of df under no unmeasured confounding.

4.4.2 Mixture of a bivariate normal distribution and a bivariate Gamma distribution

To isolate the impact of skewness, we generate data from a bivariate normal-gamma mixture distribution having the same mean and covariance as for the bivariate normal model in Section 2.1. That is,

$$\boldsymbol{\epsilon}_i = (\epsilon_{1,i}, \epsilon_{2,i}) \sim \pi \times N(\boldsymbol{\epsilon}_i; \boldsymbol{\mu}_{\boldsymbol{\epsilon}}, \boldsymbol{\Sigma}) + (1-\pi) \times G(\boldsymbol{\epsilon}_i; \boldsymbol{\mu}_{\boldsymbol{\epsilon}}, \boldsymbol{\Sigma}), \quad (13)$$

where $\boldsymbol{\mu}_{\boldsymbol{\epsilon}}$ and $\boldsymbol{\Sigma}$ are as in (12), π is the mixing proportion, and $N(\boldsymbol{\epsilon}_i; \boldsymbol{\mu}_{\boldsymbol{\epsilon}}, \boldsymbol{\Sigma})$ denotes the bivariate normal distribution and $G(\boldsymbol{\epsilon}_i; \boldsymbol{\mu}_{\boldsymbol{\epsilon}}, \boldsymbol{\Sigma})$ denotes the bivariate gamma distribution with the mean $\boldsymbol{\mu}_{\boldsymbol{\epsilon}}$ and variance $\boldsymbol{\Sigma}$. Draws from the mixture distribution are made by evaluating $\boldsymbol{\epsilon}_i = \boldsymbol{\Lambda}^T(\pi\boldsymbol{\nu}_1 + (1-\pi)\boldsymbol{\nu}_2)$, where $\boldsymbol{\Lambda}$ is the Choleski decomposition of $\boldsymbol{\Sigma}$ (i.e., $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}^T\boldsymbol{\Lambda}$), assigning the elements of $\boldsymbol{\nu}_1$ to be standard normal random variables, and

assigning the elements of $\boldsymbol{\nu}_2$ to be univariate gamma random variables with both mean and variance equal to 1. Smaller π allows greater departure from the bivariate normal distribution, generating more skewed data. We consider six settings of ρ_{GEN}=0 and 0.2 by π=0.2, 0.5, and 0.8. The other model specifications and simulation settings are as in Section 4.4.1.

As π increases, the BSEM estimator of ψ is less biased (Figure 2(d)), more precise and has better coverage (not presented). The changes in bias, precision and coverage due to skewness are much more pronounced in the presence of unmeasured confounding ($\rho_{GEN} = 0.2$) and are greater than those under no unmeasured confounding ($\rho_{GEN} = 0$).

5. DISCUSSION

In this paper, we have developed a novel Bayesian structural equations model (BSEM) to estimate the causal effect of treatment on survival subject to censoring accounting for unmeasured confounding by jointly modeling survival time and treatment using SEM. The approach assumes an underlying bivariate normal distribution for the log-survival time and propensity to be "treated." Bayesian MCMC techniques were used for estimation which, for computational efficiency, included treating the censored survival times as unknown parameters to be estimated. The methodology extends comparative effectiveness research methodology to account for both unmeasured confounding and censoring, whereas almost all prior work has focused on one problem or the other.

In the rAAA data analysis, the BSEM appeared to offer the most justifiable results among the applied methods (the AFT model, the linear model implied by the 2SLS procedure, and the BSEM) and was robust to different priors for ρ.

The BSEM performed well in finite samples with the Bayesian estimator appearing to be consistent under the model even when the prior for the unknown selection parameter was not centered on the true mean. Other simulations revealed that a stronger IV led to more robust estimation of the casual effect of treatment and that different censoring rates had little impact, implying the model made efficient use of information in the censored survival times. However, when unmeasured confounding was present, the joint model was sensitive to departures from the bivariate normal distribution in terms of the propensity for outliers and skewness.

The censoring mechanism assumed in this paper is non-informative. That is, the dependence considered in the proposed method is the one induced by the unmeasured confounders between treatment and survival time. The method can be extended to informative censoring by adding an equation for censored time which is also explained by treatment and confounders and by assuming the unmeasured confounders among the three equations for survival time, censoring time, and treatment selection to be correlated.

The parameter ρ of the bivariate normal distribution represents the net effect of unmeasured confounding and is easily interpreted. However, estimation of ρ and the other model parameters relies on the untestable bivariate normality assumption, and misspecifying the assumption can lead to substantial bias, which is evident in the simulations for the sensitivity of the BSEM to wrongly assumed distributions. An alternative involves accounting for censoring and confounding in sequence, not simultaneously. For example, first use the BSEM solely to multiply impute censored survival times. Then apply a traditional IV analysis to each completed data set. This method might allow censoring to be accounted while preserving the robustness of the IV procedure to distributional misspecifications. Therefore, the methodology derived here is not limited to the parametric domain and can also be thought of as overcoming a problem of informative censoring prior to applying IV methods for complete (non-censored) data. We are currently evaluating the just described procedure.

ACKNOWLEDGEMENTS

This study is part of the NIH-funded project (1RC4MH092717-01 NIH/NIMH) to develop novel methods for combining the strengths of RCT and observational data.

REFERENCES

Abbring, J.H. and van den Berg, G.J. (2005), "Social experiments and instrumental variables with duration outcomes," *Working Paper, IFAU - Institute for Labour Market Policy Evaluation*, No. 2005:11

Bijwaard, G.E. (2009), "Instrumental Variable Estimation for Duration Data," *Tinbergen Institute Discussion Paper*, TI 2008-032/4.

Bhattacharya, J., Goldman, D., and McCaffrey, D. (2006), "Estimating Probit Models with Self-selected Treatments," *Statistics in Medicine*, 25(3), 389–413.

Blundell, R., and Powell, J.L. (2007), "Censored regression quantiles with endogenous regressors," *Journal of Econometrics*, 141, 65–83.

Bosco, J.L.F., Silliman, R.A., Thwin, S.S., Geiger, A.M., Buist, D.S., Prout, M.N., Yood, M.U., Haque, R., Wei, F., and Lash, T.L. (2010), "A Most Stubborn Bias: No Adjustment Method Fully Resolves Confounding by Indication in Observational Studies," *Journal of Clinical Epidemiology*, 63(1), 64–74.

Brookhart, M.A., and Schneeweiss, S. (2007), "Preference-based Instrumental Variable Methods for the Estimation of Treatment Effects: Assessing Validity and Interpreting Results," *The International Journal of Biostatistics*, 3(1), 14.

Chernozhukov, V., Fernandez-Val, I., and Kowalski, K. (2011), "Quantitle Regression with Censoring and Endogeneity," *Cowels Foundation Discussion Paper*, No. 1797.

Chib, S., and Hamilton, B.H. (2000), "Bayesian analysis of cross-section and clustered data treatment models," *Journal of Econometrics*, 97, 25–50.

Chib, S., and Hamilton, B.H. (2002), "Semiparametric Bayes analysis of longitudinal data treatment models," *Journal of Econometrics*, 110, 67–89.

Copas, J.B., and Li, H.G. (1997), "Inference for Non-random Samples," *Journal Royal Statistical Society B*, 59, 55–77.

Goldman, D.P, Bhattacharya, J., McCaffrey, D.F., Duan, N., Leibowitz, A.A., Joyce, G.F., and Morton, S.C. (2001), "Effect of Insurance on Mortality in an HIV-positive Population in Care," *Journal of American Statistical Association*, 96(455), 883–894.

Gore, J.L., Litwin, M.S., Lai, J., Yano, E.M., Madison, R., Setodji, C., and Adams, J.L. (2010), "Saigal CS. Use of Redical Cystectomy for Patients With Invasive Bladder Cancer," *Journal of the National Cancer Institute*, 102(11), 802–811.

Imbens, G.W., and Angrist, J.D. (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Heckman, J.J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–960.

Hernan, M.A., Hernandez-Diaz, S., and Robins, J.M. (2004), "A Structural Approach to Selection Bias," *Epidemiology*, 15, 615–625.

O'Malley, A.J., Cotterill, P., Schermerhorn, M.L., Landon, B.E. (2011a), "Improving Observational Study Estimates of Treatment Effects Using Joint Modeling of Selection Effects and Outcomes: the Case of AAA Repair," *Medical Care*, 49(12), 1126–1132.

O'Malley, A.J., Frank, R.G., and Normand, S.L. (2011b), "Estimating Cost-offsets of New Medications: Use of New Antipsychotics and Mental Health Costs for Schizophrenia," *Statistics in Medicine*, 30(16), 1971–1988.

Palmer, C. (2013), "Why Did So Many Subprime Borrowers Default During the Crisis: Loose Credit or Plummeting Prices?" *Job Market Paper*, `http://web.mit.edu/cjpalmer/www/CPalmer_JMP.pdf/`

Terza, J., Basu, A., and Rathouz, P. (2008), "Two-stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling," *Journal of Health Economics*, 27, 531–543.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701–1728.

Yu, M., Xie, D., Wang, X., Weiner, M.G., and Tannen, R.L. (2012), "Prior Event Rate Ratio Adjustment: Numerical Studies of a Statistical Method to Address Unrecognized Confounding in Observational Studies," *Pharmacoepidemiology and Drug Safety*, 21(S2), 60–68

Zeng, F., O'Leary, J.F., Sloss, E.M., Lopez, M.S., Dhanani, N., and Melnick, G. (2006), "The Effect of Medicare Health Maintenance Organizations on Hospitalization Rates for Ambulatory Care-sensitive Conditions," *Medical Care*, 44(10), 900–907.

APPENDICES

The conditional posterior distributions discussed in Section 2.3 are provided below.

1) Posterior distribution of Y_i^* (potential survival time)

Assuming a non-informative prior, we derive the conditional posterior distribution of Y_i^*,

$$p(Y_i^*|\psi,\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\theta},\sigma_1^2,\rho,C_i) \propto \phi\big(Y_i^*;\mu_{y^*,i},\sigma_1^2\big)\Phi\big(\mu_{w|y^*,i}\big)^{W_i}\big(1-\Phi\big(\mu_{w|y^*,i}\big)\big)^{1-W_i} I(Y_i^* \geq \log(C_i),$$

where $\mu_{y^*,i} = \psi W_i + \boldsymbol{\beta}^T \boldsymbol{X}_i$ and $\mu_{w|y^*,i} = \big[\boldsymbol{\lambda}\boldsymbol{Z}_i + \boldsymbol{\theta}^T \boldsymbol{X}_i + \rho(Y_i^* - \mu_{y,i})/\sigma_1\big]/\sqrt{1-\rho^2}$ for $\rho \neq 1$. Candidate values of Y_i^* are generated using a Metropolis-Hastings (M-H) independence step (Tierney 1994). Instead of the typically-used random walk step, the candidate generating density is the fixed approximation to the true conditional posterior given by $p\big(Y_i^*;\mu_{y^*,i},\sigma_1^2\big) \propto \phi\big(Y_i^*;\mu_{y^*,i},\sigma_1^2\big) I(Y_i^* \geq \log(C_i))$.

2) Posterior distribution of $\boldsymbol{\beta}$

Under the improper uniform prior for $\boldsymbol{\beta}$ (all values of $\boldsymbol{\beta}$ considered equally likely), the conditional posterior distribution of $\boldsymbol{\beta}$ is available in closed form, allowing use of a Gibbs Sampling step to generate values of $\boldsymbol{\beta}$. It is given by

$$\boldsymbol{\beta}|\sigma_1^2,\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\theta}},\tilde{\rho} \sim N\big((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\tilde{\boldsymbol{Y}},\ \sigma_1^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}\big),$$

where $\boldsymbol{X}$ is the matrix with i-th row $\boldsymbol{X}_i^T$ and $\tilde{\boldsymbol{Y}}$ has the i-th element $\tilde{Y}_i = Y_i^a - \psi W_i$, where $Y_i^a = Y_i^{\Delta_i}(Y_i^*)^{(1-\Delta_i)}$.

3) Posterior distribution of σ_1^2

Under the Inverse Gamma conjugate prior, IG (υ_1, υ_2), σ_1^2 has the conditional posterior distribution,

$$\sigma_1^2|\boldsymbol{\beta},\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\theta}},\tilde{\rho} \sim IG\left(\upsilon_1 + \frac{n}{2},\ \upsilon_2 + \frac{1}{2}\sum_{i=1}^{n}\big(\tilde{Y}_i - \boldsymbol{\beta}^T\boldsymbol{X}_i\big)^2\right)$$

We choose υ_1(=0.001) and υ_2(=0.001) so that the prior is diffuse with large variance (= ∞).

4) Posterior distribution of ($\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\theta}},\tilde{\rho}$)

Assuming non-informative priors for the two transformed parameters $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\boldsymbol{\theta}}$ and the informative prior (including a diffuse prior as a special case) given in (10) for $\tilde{\rho}$, we obtain the joint conditional posterior distribution for $(\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\theta}},\tilde{\rho})$ given by

$$p(\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\theta}},\tilde{\rho}|\boldsymbol{\beta},\sigma_1^2,\psi)$$
$$\propto \prod_{i=1}^{n}\Phi\big(\mu_{w|\tilde{y},i}\big)^{W_i}\big(1-\Phi\big(\mu_{w|\tilde{y},i}\big)\big)^{1-W_i}\frac{\sigma_1}{\sqrt{1+\tilde{\rho}^2\sigma_1^2}}\left(1+\frac{\tilde{\rho}\sigma_1}{\sqrt{1+\tilde{\rho}^2\sigma_1^2}}\right)^{\nu_1-1}\left(1-\frac{\tilde{\rho}\sigma_1}{\sqrt{1+\tilde{\rho}^2\sigma_1^2}}\right)^{\nu_2-1}.$$

The covariance parameter used in the vector-normal distribution for updating $(\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\theta}},\tilde{\rho})$ is the covariance of the maximum likelihood estimator of $(\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\theta}},\tilde{\rho})$ under a normal PDF based on only the non-censored data. After generating each new candidate value, the values of $\tilde{\boldsymbol{\lambda}}$, $\tilde{\boldsymbol{\theta}}$, and $\tilde{\rho}$ are transformed back to the original parameters $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$, and ρ for use in the other steps of the MCMC algorithm.

5) Posterior distribution of ψ

Under an improper flat prior for ψ, the conditional posterior distribution of ψ is given by

$$p(\psi|\boldsymbol{\beta},\sigma_1^2\tilde{\boldsymbol{\lambda}},\tilde{\boldsymbol{\theta}},\tilde{\rho}) \propto \phi\big(Y_i^a;\mu_{y^a,i},\sigma_1^2\big)\Phi\big(\mu_{w|y^a,i}\big)^{W_i}\big(1-\Phi\big(\mu_{w|y^a,i}\big)\big)^{1-W_i},$$

where $Y_i^a = Y_i^{\Delta_i}(Y_i^*)^{(1-\Delta_i)}$, and $\mu_{y^a,i}$ and $\mu_{w|y^a,i}$ have Y_i^a in place of Y_i in (6). A M-H step with a random-walk normal distribution as the candidate generating function is used for updating ψ.