

# Efficient Estimation for Semiparametric Structural Equation Models With Censored Data

Kin Yau Wong, Donglin Zeng, and D. Y. Lin

Department of Biostatistics, University of North Carolina, Chapel Hill, NC

## ABSTRACT

Structural equation modeling is commonly used to capture complex structures of relationships among multiple variables, both latent and observed. We propose a general class of structural equation models with a semiparametric component for potentially censored survival times. We consider nonparametric maximum likelihood estimation and devise a combined expectation-maximization and Newton-Raphson algorithm for its implementation. We establish conditions for model identifiability and prove the consistency, asymptotic normality, and semiparametric efficiency of the estimators. Finally, we demonstrate the satisfactory performance of the proposed methods through simulation studies and provide an application to a motivating cancer study that contains a variety of genomic variables. Supplementary materials for this article are available online.

## ARTICLE HISTORY

Received April 2016  
Revised January 2017

## KEYWORDS

Integrative analysis; Joint modeling; Latent variables; Model identifiability; Nonparametric maximum likelihood estimation; Survival analysis

## 1. Introduction

Structural equation modeling (SEM) is a very general and powerful approach to capture complex relationships among multiple factors, both observed and latent (Bollen 1989). A typical SEM framework consists of a structural model that connects latent variables and a measurement model that relates latent variables to observed variables. SEM is extremely popular in the social sciences and psychology, where unmeasured quantities and psychological constructs, such as human intelligence and creativity, can be related to and investigated through observed data. The text of Bollen (1989) has been cited more than 20,000 times. Recently, SEM has gained popularity in medical and public health research (Dahly, Adair, and Bollen 2009; Naliboff et al. 2012).

Our interest in SEM was motivated by its potential application to integrative analysis in genomic studies. Recent technological advances have made it possible to collect different types of genomic data, including DNA copy number, SNP genotype, DNA methylation level, and expression levels of mRNA, microRNA, and protein, on a large number of subjects. There is a growing interest in integrating these genomic platforms so as to understand their biological relationships and predict disease progression and death, which are considered potentially censored survival times (The Cancer Genome Atlas (TCGA); <https://tcga-data.nci.nih.gov/tcga/>).

SEM with discrete survival times has been studied by Rabe-Hesketh, Yang, and Pickles (2001), Rabe-Hesketh, Skroindal, and Pickles (2004), Muthén and Masyn (2005), and Moustaki and Steele (2005). For continuous survival time, Larsen (2004, 2005) adopted the proportional hazards model (Cox 1972) with a single latent variable to capture the association between the survival time and other observed variables;

Asparouhov, Masyn, and Muthén (2006) considered a more general formulation of the association among the latent and observed variables. SEM with the Cox proportional hazards model for the survival component has been adopted for more complex settings, such as multivariate survival times (Stoolmiller and Snyder 2006) and competing risks (Stoolmiller and Snyder 2014). A popular software program, Mplus (Muthén and Muthén 1998–2015), has implemented SEM with survival data under the proportional hazards model. The estimation of the nonparametric baseline hazard function is based on piecewise-constant splines, and no theoretical justification is available. In fact, the standard error estimator for the baseline hazard function is incorrect.

In this article, we propose a general SEM framework that includes a semiparametric component of the measurement model for potentially censored survival times. Specifically, we formulate the effects of latent and observed covariates on survival times through a broad class of semiparametric transformation models that includes the proportional hazards model as a special case. The observed covariates may include manifest variables that depend on latent variables. We study nonparametric maximum likelihood estimation (NPMLE), under which the cumulative hazard functions are estimated by step functions with jumps at observed survival times.

The proposed SEM is reminiscent of joint modeling for survival and longitudinal data (Henderson, Diggle, and Dobson 2000; Tsiatis and Davidian 2004). With the latter, the observed longitudinal variables are considered error-prone measurements of some underlying latent variables, but the measurements themselves are not causal determinants of the survival time. By contrast, our SEM framework allows latent variables to have direct effects on survival times, as well as indirect effects

through other manifest variables. In addition, our framework accommodates much more complex relationships among latent variables.

A major challenge in our theoretical development is model identifiability. Even for an SEM with normally distributed variables, no single set of conditions exists that is both necessary and sufficient for model identifiability. Methods that deal with special cases of the normal SEM were proposed by Bollen (1989), Reilly and O'Brien (1996), Vicard (2000), and Bollen and Davis (2009), among others. Most of the methods are based on the fact that identifiability can be established by solving the equations relating the first two model-implied moments to the sample moments. This approach is not directly applicable to models with nonparametric components, as infinite-dimensional parameters cannot be identified through a finite number of equations. Because the proportional hazards structure results in a likelihood function that takes the form of a Laplace transform, however, we are able to develop sufficient conditions under which the identifiability of a semiparametric SEM can be established by inspecting simpler parametric models.

Another theoretical challenge is the invertibility of the information operator. For the information operator to be invertible, we require that the score statistic along any nontrivial submodel is nonzero. As in the case of model identifiability, general conditions for the invertibility of the information operator for semiparametric models do not exist. In the existing work involving latent variables for survival times (Kosorok, Lee, and Fine 2004; Zeng and Lin 2010), verifying the invertibility of the information operator involves inspecting the local behavior of the score statistic around the zero survival time. This approach does not make full use of the variability of the score statistic contributed by the survival times and cannot deal with the proposed general modeling framework. We show that the invertibility of the information operator can be verified by inspecting the parametric components of the SEM under some mild conditions in the survival model.

The rest of this article is structured as follows. In Section 2, we formulate the model and describe our approach to establish model identifiability. In Section 3, we discuss the numerical implementation of the NPMLE. In Section 4, we present theoretical results for model identifiability and describe the asymptotic properties of the estimators. In Section 5, we report the results from simulation studies. In Section 6, we provide an application to the TCGA data, which motivated this work. We make some concluding remarks in Section 7 and relegate theoretical proofs to the Appendix.

## 2. Basic Framework

### 2.1. Model and Likelihood

Let  $\boldsymbol{\eta}$  denote a  $q$ -vector of latent variables,  $\mathbf{Y}$  denote an  $r$ -vector of uncensored manifest variables,  $(T_1, \dots, T_K)$  denote  $K$  potentially censored survival times, and  $\mathbf{W}$  and  $\mathbf{Z}$  denote two vectors of observed covariates. Without loss of generality, assume that the support of the covariates includes zero. We specify the conditional distributions of  $\boldsymbol{\eta}$  given  $\mathbf{Z}$ ,  $\mathbf{Y}$  given  $\mathbf{Z}$  and  $\boldsymbol{\eta}$ , and  $T_k$  given

$\mathbf{W}$ ,  $\mathbf{Z}$ ,  $\mathbf{Y}$ , and  $\boldsymbol{\eta}$  as follows:

$$\boldsymbol{\eta} \mid \mathbf{Z} \sim F_{\boldsymbol{\eta}}(\cdot \mid \mathbf{Z}; \mathbf{v}), \quad (1)$$

$$\mathbf{Y} \mid (\mathbf{Z}, \boldsymbol{\eta}) \sim F_Y(\cdot \mid \mathbf{Z}, \boldsymbol{\eta}; \boldsymbol{\psi}), \quad (2)$$

$$\begin{aligned} \Lambda_{T_k}(t \mid \mathbf{W}, \mathbf{Z}, \mathbf{Y}, \boldsymbol{\eta}) \\ = G_k \{ \Lambda_k(t) e^{\mathbf{W}^T \boldsymbol{\vartheta}_k + \mathbf{Z}^T \boldsymbol{\beta}_k + \mathbf{Y}^T \boldsymbol{\alpha}_k + \boldsymbol{\eta}^T \boldsymbol{\phi}_k} \}, \quad k = 1, \dots, K, \end{aligned} \quad (3)$$

where  $F_{\boldsymbol{\eta}}(\cdot \mid \mathbf{Z}, \mathbf{v})$  denotes a  $q$ -variate normal distribution function indexed by a parameter vector  $\mathbf{v}$ ,  $F_Y(\cdot \mid \mathbf{Z}, \boldsymbol{\eta}; \boldsymbol{\psi})$  denotes an  $r$ -variate parametric distribution function indexed by a parameter vector  $\boldsymbol{\psi}$ ,  $\Lambda_{T_k}$  is the cumulative hazard function of  $T_k$  given  $(\mathbf{W}, \mathbf{Z}, \mathbf{Y}, \boldsymbol{\eta})$ ,  $G_k$  is a known increasing function,  $\Lambda_k$  is an unspecified positive increasing function with  $\Lambda_k(0) = 0$ , and  $(\boldsymbol{\vartheta}_k, \boldsymbol{\beta}_k, \boldsymbol{\alpha}_k, \boldsymbol{\phi}_k)$  are unknown regression parameters.

Model (1) is the structural model of the latent variables. Model (2) is the measurement model of  $\mathbf{Y}$ . We assume that  $\mathbf{Y}$  and  $\boldsymbol{\eta}$  are independent of  $\mathbf{W}$  given  $\mathbf{Z}$ . Models (1) and (2) represent the existing SEM framework with  $\mathbf{Y}$  not restricted to be normally distributed. Equation (3) includes the proportional hazards and proportional odds models as special cases with the choices of  $G_k(x) = x$  and  $G_k(x) = \log(1 + x)$ , respectively. The proportional hazards model has been considered in the literature.

The survival time  $T_k$  is subject to right censoring by  $C_k$ . It is assumed that  $(C_1, \dots, C_K)$  are independent of  $(T_1, \dots, T_K)$  and  $\boldsymbol{\eta}$  conditional on  $\mathbf{Y}$ ,  $\mathbf{Z}$ , and  $\mathbf{W}$ . Define  $\tilde{T}_k = \min(T_k, C_k)$  and  $\Delta_k = I(T_k \leq C_k)$ , where  $I(\cdot)$  is the indicator function. For a sample of size  $n$ , the observed data consist of  $\mathcal{O}_i \equiv (\tilde{T}_{1i}, \dots, \tilde{T}_{Ki}, \Delta_{1i}, \dots, \Delta_{Ki}, \mathbf{Y}_i, \mathbf{Z}_i, \mathbf{W}_i)$  ( $i = 1, \dots, n$ ).

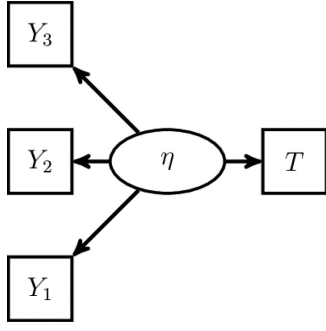
Let  $\boldsymbol{\theta}$  denote the collection of all Euclidean parameters, and write  $\mathcal{A} = (\Lambda_1, \dots, \Lambda_K)$ . The likelihood function for  $\boldsymbol{\theta}$  and  $\mathcal{A}$  is proportional to

$$\begin{aligned} L_n(\boldsymbol{\theta}, \mathcal{A}) = \prod_{i=1}^n \int \prod_{k=1}^K [\lambda_k(\tilde{T}_{ki}) e^{\mathbf{W}_i^T \boldsymbol{\vartheta}_k + \mathbf{Z}_i^T \boldsymbol{\beta}_k + \mathbf{Y}_i^T \boldsymbol{\alpha}_k + \boldsymbol{\eta}^T \boldsymbol{\phi}_k} \\ \times G'_k \{ \Lambda_k(\tilde{T}_{ki}) e^{\mathbf{W}_i^T \boldsymbol{\vartheta}_k + \mathbf{Z}_i^T \boldsymbol{\beta}_k + \mathbf{Y}_i^T \boldsymbol{\alpha}_k + \boldsymbol{\eta}^T \boldsymbol{\phi}_k} \}]^{\Delta_{ki}} \\ \times \exp \left[ - G_k \{ \Lambda_k(\tilde{T}_{ki}) e^{\mathbf{W}_i^T \boldsymbol{\vartheta}_k + \mathbf{Z}_i^T \boldsymbol{\beta}_k + \mathbf{Y}_i^T \boldsymbol{\alpha}_k + \boldsymbol{\eta}^T \boldsymbol{\phi}_k} \} \right] \\ \times f_Y(\mathbf{Y}_i \mid \mathbf{Z}_i, \boldsymbol{\eta}; \boldsymbol{\psi}) f_{\boldsymbol{\eta}}(\boldsymbol{\eta} \mid \mathbf{Z}_i; \mathbf{v}) d\boldsymbol{\eta}, \end{aligned} \quad (4)$$

where  $f'(x) = df(x)/dx$  for any function  $f$ ,  $\lambda_k = \Lambda'_k$ ,  $f_Y = F'_Y$ , and  $f_{\boldsymbol{\eta}} = F'_{\boldsymbol{\eta}}$ . The NPMLE is defined to be the maximizer of  $L_n(\boldsymbol{\theta}, \mathcal{A})$ , in which  $\Lambda_k$  is treated as a step function with jumps at  $\tilde{T}_{ki}$  with  $\Delta_{ki} = 1$  ( $i = 1, \dots, n$ ).

### 2.2. Model Identifiability

We describe our approach to establish model identifiability in this section and defer the technical details to Section 4. The identifiability results can be summarized by two simple rules. Suppose that we have arranged the survival times such that for some  $0 \leq K_1 \leq \min(q, K)$ , each of  $(T_1, \dots, T_{K_1})$  regresses on and only on one latent variable and a set of covariates that are independent of the latent variables. (We allow  $K_1 = 0$  if no survival time satisfies the given conditions, in which case Rule 1 below is vacuous.) We call an observed variable  $X$  an indicator of a latent variable  $\eta$  if  $X$  follows a generalized linear model with  $\eta$  as a covariate and is independent of all other manifest variables and survival times conditional on  $\eta$ . We have the following rules:



**Figure 1.** The first example of SEM to illustrate the identifiability rules. The SEM consists of one latent variable, one survival time, and three conditionally independent normal manifest variables.

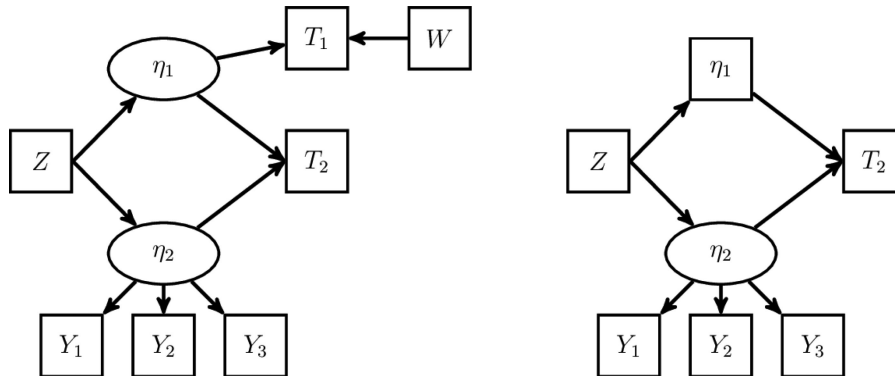
1. The latent variables attached to  $(T_1, \dots, T_{K_1})$  can be treated as observed if each of  $(T_1, \dots, T_{K_1})$  depends on at least one observed covariate.
2. If each latent variable has a separate continuous indicator and the distributions of the latent variables and the indicators are identifiable, then the whole model is identifiable.

To illustrate the usefulness of the two identifiability rules, we present two examples.

*Example 1.* Consider the model depicted in Figure 1. In the model,  $Y_1$ ,  $Y_2$ , and  $Y_3$  are conditionally independent normal manifest variables of  $\eta$ , and  $T$  is a survival time that follows the proportional hazards model with covariate  $\eta$ . Assume that the regression parameter of  $Y_1$  on  $\eta$  is fixed to be one,  $E(\eta) = 0$ , and the regression parameters of  $\eta$  in the models of  $Y_2$  and  $Y_3$  are nonzero.

The above model is similar to the joint model for survival and longitudinal variables. By Bollen (1989)'s three-indicator rule, the model of  $(Y_1, Y_2, Y_3, \eta)$  is identifiable. With  $Y_1$  serving as an indicator of  $\eta$ , Rule 2 implies that the remaining parameters are identifiable. Note that Rule 1 is not applicable in this case because  $T$  does not depend on an independent covariate. In fact, the model is not identifiable without  $(Y_1, Y_2, Y_3)$  because the scale of the baseline hazard function and the variance of  $\eta$  cannot be separated.

*Example 2.* Consider the model depicted on the left-hand side of Figure 2. In the model,  $Y_1$ ,  $Y_2$ , and  $Y_3$  are conditionally independent normal manifest variables of  $\eta_2$ ,  $T_1$  is a survival time that follows the proportional hazards model with covariates  $W$  and



**Figure 2.** The second example of SEM to illustrate the identifiability rules. The left panel is an SEM that consists of two latent variables, two observed covariates, two survival times, and three conditionally independent normal manifest variables. The right panel is an intermediate step in identifying the SEM on the left.

$\eta_1$ , and  $T_2$  is a survival time that follows the proportional hazards model with covariates  $\eta_1$  and  $\eta_2$ . Assume that  $W$  and  $Z$  are nonconstant and linearly independent, the regression parameters for the latent variables in the models of  $T_1$  and  $Y_1$  are fixed to be one,  $E(\eta_1) = E(\eta_2) = 0$ , and the regression parameters of  $W$  in the model of  $T_1$  and  $\eta_2$  in the models of  $Y_2$  and  $Y_3$  are nonzero.

First, we use  $T_1$  to help identify the latent variable distributions. By Rule 1,  $\eta_1$  can be treated as observed when identifying the model. The problem thus reduces to identifying the model shown on the right-hand side of Figure 2. The model can then be shown identifiable by the arguments used in Example 1.

### 3. Computation of the NPMLE

In this section, we use  $\mathbf{Z}$  to denote both  $W$  and  $Z$  with  $\beta_k$  ( $k = 1, \dots, K$ ) as the corresponding vector of regression parameters. Application of a transformation  $G_k$  can be viewed as inclusion of an extra latent variable  $\log s_k$  in the regression equation, where  $s_k$  is a random variable with density  $g_k$  such that  $G_k(x) = -\log \int_0^\infty e^{-xt} g_k(t) dt$ . We adopt the expectation-maximization (EM) algorithm (Dempster, Laird, and Rubin 1977) by treating the latent variables, including those introduced by the transformations, as missing data. We perform occasional Newton-Raphson steps to speed up the convergence.

In the combined algorithm, either an EM step or a Newton-Raphson step is performed at each iteration. To avoid confusion, we call the latter an outer Newton-Raphson step. For an EM step, note that the conditional expectation for any function  $\varphi$  of  $(\eta_i, s_i) \equiv (\eta_i, s_{1i}, \dots, s_{K_i})$  given the observed data is

$$\begin{aligned}
 & E\{\varphi(\eta_i, s_i) \mid \mathcal{O}_i\} \\
 &= \mathcal{C}^{-1} \int \int \varphi(\boldsymbol{\eta}, \mathbf{s}) \prod_{k=1}^K \left( \left[ \Lambda_k\{\tilde{T}_{ki}\} s_k e^{\mathbf{Z}_i^T \boldsymbol{\beta}_k + \mathbf{Y}_i^T \boldsymbol{\alpha}_k + \eta^T \boldsymbol{\phi}_k} \right]^{\Delta_{ki}} \right. \\
 & \quad \times \exp \left( - \int_0^{\tilde{T}_{ki}} s_k e^{\mathbf{Z}_i^T \boldsymbol{\beta}_k + \mathbf{Y}_i^T \boldsymbol{\alpha}_k + \eta^T \boldsymbol{\phi}_k} d\Lambda_k(t) \right) \Bigg) \\
 & \quad \times f_Y(\mathbf{Y}_i \mid \mathbf{Z}_i, \boldsymbol{\eta}; \boldsymbol{\psi}) f_\eta(\boldsymbol{\eta} \mid \mathbf{Z}_i; \boldsymbol{\nu}) g(\mathbf{s}) d\boldsymbol{\eta} ds_1 \cdots ds_K,
 \end{aligned}$$

where  $\Lambda_k\{t\}$  is the jump size of the step function  $\Lambda_k$  at  $t$ ,  $\mathbf{s} = (s_1, \dots, s_K)$ ,  $g(\mathbf{s}) = \prod_{k=1}^K g_k(s_k)$ , and  $\mathcal{C}$  equals the above integral evaluated at  $\varphi(\cdot, \cdot) = 1$ . We use the Gauss-Hermite quadrature to approximate the integrals. To reduce the number of abscissas, we adopt an adaptive quadrature approach (Liu

and Pierce 1994). Denote the approximation of the conditional expectation as  $\hat{E}(\cdot)$ . After taking expectation on the functions involved, we update  $(\beta_k, \alpha_k, \phi_k)$  by the one-step Newton-Raphson algorithm on

$$\sum_{i=1}^n \log \left[ \frac{\hat{E} \{ s_{ki} \exp(\mathbf{Z}_i^T \beta_k + \mathbf{Y}_i^T \alpha_k + \eta_i^T \phi_k) \}}{\sum_{j=1}^n I(\tilde{T}_{kj} \geq \tilde{T}_{ki}) \hat{E} \{ s_{kj} \exp(\mathbf{Z}_j^T \beta_k + \mathbf{Y}_j^T \alpha_k + \eta_j^T \phi_k) \}} \right]^{\Delta_{ki}}$$

Then, we update the cumulative baseline hazard function by

$$\hat{\Lambda}_k\{T_{ki}\} = \frac{\Delta_{ki}}{\sum_{j=1}^n I(\tilde{T}_{kj} \geq \tilde{T}_{ki}) \hat{E} \{ s_{kj} \exp(\mathbf{Z}_j^T \beta_k + \mathbf{Y}_j^T \alpha_k + \eta_j^T \phi_k) \}},$$

where  $(\beta_k, \alpha_k, \phi_k)$  are evaluated at the current estimates. In addition, we update the remaining parameters at the maximum of

$$\sum_{i=1}^n \hat{E}[\log\{f_Y(Y_i | \mathbf{Z}_i, \boldsymbol{\eta}_i; \boldsymbol{\psi}) f_{\eta}(\boldsymbol{\eta}_i | \mathbf{Z}_i; \mathbf{v})\}].$$

If a closed-form solution is not available, then we apply the one-step Newton-Raphson algorithm to the above expression instead. The above algorithm can be generalized to the case where components of  $\mathbf{Y}_i$  that do not appear in the model of the survival times are missing at random for some subjects. In this case, we simply drop the corresponding  $f_Y$  terms in the evaluation of  $\hat{E}$  and the complete-data log-likelihood.

For an outer Newton-Raphson step, we apply the one-step Newton-Raphson algorithm directly to the logarithm of  $L_n(\boldsymbol{\theta}, \mathcal{A})$  given in (4) using a similar adaptive quadrature approximation. At the current estimates, the first derivative of  $\log L_n(\boldsymbol{\theta}, \mathcal{A})$ , that is, the score statistic, is the same as the first derivative of the expected complete-data log-likelihood. The Hessian matrix used in the Newton-Raphson algorithm can be obtained by Louis's (1982) formula.

To determine whether an EM step or an outer Newton-Raphson step is to be performed, we keep track of the difference in the log-likelihood at the previous iteration, either an EM or an outer Newton-Raphson step, and the difference at the previous outer Newton-Raphson step. For each iteration, if the log-likelihood difference at the previous step is too small relative to that at the previous outer Newton-Raphson step, then an outer Newton-Raphson step is performed; otherwise, an EM step is performed. Upon convergence, Louis's (1982) formula is used to obtain the information matrix for the estimation of the standard errors.

The reason that we use a combination of the EM and Newton-Raphson algorithms instead of the Newton-Raphson algorithm alone is two-fold. First, EM steps are more stable, which is important, especially in early iterations. Second, in the estimation of the survival model under the EM algorithm, the regression parameters can be obtained by maximizing the partial-likelihood-type function, and the estimators of the baseline hazard functions take the form of the Breslow estimator. Unlike the Newton-Raphson algorithm, the EM algorithm does not involve the inversion of a high-dimensional matrix.

## 4. Theoretical Properties

### 4.1. Identifiability Conditions

As discussed in Section 2, we set aside  $K_1$  survival times,  $T_1, \dots, T_{K_1}$ , that are used to identify the distribution of the underlying latent variables. We assume that  $\text{span}(\phi_1, \dots, \phi_{K_1}) = \mathbb{R}^{K_1}$ . We can choose the  $K_1$  survival times such that each is associated with a few, preferably only one, latent variables. ( $K_1$  is allowed to be 0, in which case we rely solely on the manifest variable  $\mathbf{Y}$  to identify the latent variable distribution.) Without loss of generality, we assume that  $\phi_k = \mathbf{e}_k$  ( $k = 1, \dots, K_1$ ), where  $\mathbf{e}_k$  is a  $q$ -vector with 1 at the  $k$ th position and 0 elsewhere. This assumption can be satisfied by applying a linear transformation to the latent variables. Effectively, we fix the scale of the first  $K_1$  latent variables, as is common when establishing model identifiability for SEM. We partition  $\boldsymbol{\eta}$  into  $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ , where  $\boldsymbol{\eta}_1 \equiv (\eta_{11}, \dots, \eta_{1K_1})$  consists of the first  $K_1$  components of  $\boldsymbol{\eta}$ .

We consider the following identifiability conditions. For the "baseline" hazard functions  $(\Lambda_1, \dots, \Lambda_K)$ , we only require identifiability on  $[0, \tau]$ , where  $\tau$  denotes the study duration.

- (C1) If  $(\mathbf{1}, \mathbf{W}^T, \mathbf{Z}^T, \mathbf{Y}^T)^T \mathbf{c} = 0$  almost surely for some vector  $\mathbf{c}$  of appropriate dimension, then  $\mathbf{c} = \mathbf{0}$ . For  $k = 1, \dots, K$ ,  $\lambda_k$  is continuous and strictly positive on  $[0, \tau]$ , and there exists a positive and measurable function  $g_k$  such that  $\exp\{-G_k(t)\} = \int_0^\infty e^{-ts} g_k(s) dm_k(s)$ , where  $m_k$  is the Lebesgue measure or the counting measure at 1.
- (C2) For  $k = 1, \dots, K_1$ ,  $E(\mathbf{Y}^T \boldsymbol{\alpha}_k | \mathbf{Z} = \mathbf{0}) = \mathbf{0}$ , and  $E(\boldsymbol{\eta} | \mathbf{Z} = \mathbf{0}) = \mathbf{0}$ . Also, for any vectors  $\mathbf{c}_1$  and  $\mathbf{c}_2$  of appropriate dimensions,  $E(e^{\mathbf{Y}^T \mathbf{c}_1 + \boldsymbol{\eta}^T \mathbf{c}_2} | \mathbf{Z})$  is finite almost surely.
- (C3) For  $k = 1, \dots, K_1$ ,  $\boldsymbol{\vartheta}_k$  is nonzero, and  $\boldsymbol{\alpha}_k$  and  $\beta_k$  are zero.
- (C4) Consider two sets of parameters  $(\boldsymbol{\psi}, \mathbf{v})$  and  $(\tilde{\boldsymbol{\psi}}, \tilde{\mathbf{v}})$ . Let  $f_{Y, \boldsymbol{\eta}_1}$  be the density of  $(\mathbf{Y}, \boldsymbol{\eta}_1)$  given  $\mathbf{Z}$ . Then,  $f_{Y, \boldsymbol{\eta}_1}(\mathbf{Y}, \boldsymbol{\eta}_1 | \mathbf{Z}; \boldsymbol{\psi}, \mathbf{v}) = f_{Y, \boldsymbol{\eta}_1}(\mathbf{Y}, \boldsymbol{\eta}_1 | \mathbf{Z}; \tilde{\boldsymbol{\psi}}, \tilde{\mathbf{v}})$  for all  $\mathbf{Z}, \mathbf{Y}$ , and  $\boldsymbol{\eta}_1$  implies that  $\boldsymbol{\psi} = \tilde{\boldsymbol{\psi}}$  and  $\mathbf{v} = \tilde{\mathbf{v}}$ .
- (C5) Let  $(\mathbf{Y}^+, \boldsymbol{\eta}_1^+)$  be the components of  $(\mathbf{Y}, \boldsymbol{\eta}_1)$  that appear in the regression of  $T_k$  for some  $k = K_1 + 1, \dots, K$ , and let  $(\mathbf{Y}^-, \boldsymbol{\eta}_1^-)$  be the remaining components. Let  $F_{\boldsymbol{\eta}_2 | \mathbf{Y}, \boldsymbol{\eta}_1}$  be the distribution function of  $\boldsymbol{\eta}_2$  given  $(\mathbf{Z}, \mathbf{Y}, \boldsymbol{\eta}_1)$  with  $(\mathbf{Y}^-, \boldsymbol{\eta}_1^-)$  treated as a parameter vector. Then,  $\boldsymbol{\eta}_2$  is complete sufficient in  $\{F_{\boldsymbol{\eta}_2 | \mathbf{Y}, \boldsymbol{\eta}_1}(\cdot | \mathbf{Z}, \mathbf{Y}, \boldsymbol{\eta}_1) : \mathbf{Z} = \mathbf{z}_0, \mathbf{Y}^+ = \mathbf{y}_0, \boldsymbol{\eta}_1^+ = \boldsymbol{\eta}_{10}\}$  for any fixed  $\mathbf{z}_0, \mathbf{y}_0$ , and  $\boldsymbol{\eta}_{10}$ .

*Remark 1.* Condition (C1) pertains to basic requirements on the covariates, the baseline hazard functions, and the transformation functions such that the survival model with observed covariates is identifiable. If  $m_k$  is a point mass at 1, then  $G_k$  is simply the identity function. Condition (C2) fixes the location parameters of the latent variables and the manifest variables that appear in the regression models of the first  $K_1$  survival times. Condition (C3) requires that the first  $K_1$  survival times depend only on their corresponding latent variable and  $\mathbf{W}$ . The presence of a covariate besides the latent variable is necessary for distinguishing the contributions of the baseline hazard function and the latent variable to the distribution of a survival time that follows a mixture distribution. Condition (C4) requires that the model with observed  $(\mathbf{Y}, \boldsymbol{\eta}_1)$  is identifiable. Condition (C5) requires that  $\boldsymbol{\eta}_2$  is complete sufficient conditional on  $(\mathbf{Y}, \boldsymbol{\eta}_1)$ ,



where components of  $(\mathbf{Y}, \boldsymbol{\eta}_1)$  that do not appear in the regression of  $T_k$  ( $k = K_1 + 1, \dots, K$ ) are treated as parameters, and the rest are held fixed. Conditions (C2) and (C3) are vacuous if  $K_1 = 0$ , and condition (C5) is vacuous if  $K_1 = K$ .

We have the following identifiability result.

*Theorem 1.* Under Conditions (C1)–(C5), the model specified by (1)–(3) is identifiable.

*Remark 2.* The condition that  $\boldsymbol{\alpha}_k$  and  $\boldsymbol{\beta}_k$  are zero for  $k = 1, \dots, K_1$  separates the first  $K_1$  survival times from the remaining observed variables that are associated with the latent variables. This condition is used to simplify the presentation of the identifiability conditions. In the proof of [Theorem 1](#), we consider generalized versions of conditions (C3)–(C5), where  $\boldsymbol{\alpha}_k$  and  $\boldsymbol{\beta}_k$  are allowed to be nonzero.

*Remark 3.* [Theorem 1](#) implies that the distribution of the latent variable underlying a given survival time can be completely identified if the survival time only regresses on the latent variable and a set of independent covariates. Thus, the survival times make it easy to identify the model, as only a single survival time is enough to identify an underlying latent variable. By contrast, this property does not hold for normal random variables.

*Remark 4.* The derivation of model identifiability from condition (C5) uses the property of complete sufficient statistics. The derivation is applicable to general latent variable models; the general result is given by Lemma 1 in the [Appendix](#). Lemma 1 allows for the establishment of model identifiability by inspecting just a part of the model. It includes Reilly and O'Brien's (1996) side-by-side rule, which states that the loadings of an observed variable on any number of latent variables are identifiable if each of the latent variables is attached to a separate independent observed variable whose distribution is identifiable, as a special case.

## 4.2. Asymptotic Properties

Let  $d$  be the dimension of  $\boldsymbol{\theta}$ ,  $\boldsymbol{\theta}_0$  denote the true value of  $\boldsymbol{\theta}$ , and  $\Lambda_{0k}$  denote the true value of  $\Lambda_k$  ( $k = 1, \dots, K$ ). We impose the following conditions.

- (D1) The parameter  $\boldsymbol{\theta}_0$  lies in the interior of a compact set  $\Theta \subset \mathbb{R}^d$ , and the function  $\Lambda_{0k}$  is continuously differentiable with  $\lambda_{0k}(t) \equiv \Lambda'_{0k}(t) > 0$  on  $[0, \tau]$  for each  $k = 1, \dots, K$ .
- (D2) With probability one,  $P(\tilde{T}_{ki} = \tau \mid \mathbf{W}, \mathbf{Z}) > \delta_0$  ( $k = 1, \dots, K$ ) for some fixed  $\delta_0 > 0$ .
- (D3) Consider any fixed  $\mathbf{Z}$  and  $(\boldsymbol{\psi}, \mathbf{v}) \in \Theta_{\boldsymbol{\psi}, \mathbf{v}}$ , where  $\Theta_{\boldsymbol{\psi}, \mathbf{v}}$  consists of the  $(\boldsymbol{\psi}, \mathbf{v})$ -component of every  $\boldsymbol{\theta} \in \Theta$ . For any constant  $a_1 > 0$  and  $\delta = 0, 1$ ,

$$E \left\{ \int e^{a_1(1+|\mathbf{Y}|+|\boldsymbol{\eta}|)} f_Y(\mathbf{Y} \mid \mathbf{Z}, \boldsymbol{\eta}; \boldsymbol{\psi})^\delta f_\eta(\boldsymbol{\eta} \mid \mathbf{Z}; \mathbf{v}) d\boldsymbol{\eta} \right\} < \infty.$$

Also, for  $j = 1, 2, 3$ , there exists a constant  $a_2 > 0$  such that

$$\left| \frac{\frac{\partial^j}{\partial \boldsymbol{\psi}^j} f_Y(\mathbf{Y} \mid \mathbf{Z}, \boldsymbol{\eta}; \boldsymbol{\psi})}{f_Y(\mathbf{Y} \mid \mathbf{Z}, \boldsymbol{\eta}; \boldsymbol{\psi})} \right| + \left| \frac{\frac{\partial^j}{\partial \mathbf{v}^j} f_\eta(\boldsymbol{\eta} \mid \mathbf{Z}; \mathbf{v})}{f_\eta(\boldsymbol{\eta} \mid \mathbf{Z}; \mathbf{v})} \right| \leq e^{a_2(1+|\mathbf{Y}|+|\boldsymbol{\eta}|)}.$$

In addition, for some positive constants  $M_j$  and  $c_j$ ,  $\mathbf{N}_j \in \mathbb{R}^r$ , and  $\varphi_1 \in \ell^\infty(\mathbb{R}^r)$ ,

$$e^{\sum_{j=1}^q -M_j |b_j|} \leq e^{\sum_{j=1}^q (\mathbf{N}_j^\top \mathbf{Y} b_j + c_j b_j^2)} \varphi_1(\mathbf{Y}) \times f_Y(\mathbf{Y} \mid \mathbf{Z}, \boldsymbol{\eta}; \boldsymbol{\psi}) f_\eta(\boldsymbol{\eta} \mid \mathbf{Z}; \mathbf{v}) \leq e^{\sum_{j=1}^q M_j |b_j|},$$

where  $\mathbf{b} = S(\boldsymbol{\eta})$  for some one-to-one linear transformation  $S$ .

- (D4) The function  $G_k$  is four-times differentiable,  $G_k(0) = 0$ , and  $K_{1k}(1+x)^{-\kappa_{1k}} \leq G'_k(x) \leq K_{2k}(1+x)^{\kappa_{2k}}$  for some positive constants  $\kappa_{1k}$ ,  $\kappa_{2k}$ ,  $K_{1k}$ , and  $K_{2k}$ . Also,  $G_k(x)/x^{\rho_k} \rightarrow M_k$  or  $G_k(x)/\log(x) \rightarrow M_k$  as  $x \rightarrow \infty$  for some positive constants  $M_k$  and  $\rho_k$ . In addition,  $\exp\{-G_k(x)\} \leq \mu_k(1+x)^{-\kappa_{3k}}$  for some  $\mu_k$  and  $\kappa_{3k} > \kappa_{2k} + 1$ . Furthermore, for some  $r_k$ ,

$$\sup_{x \geq 0} \frac{|G''_k(x)| + |G_k^{(3)}(x)| + |G_k^{(4)}(x)|}{G'_k(x)(1+x)^{r_k}} < \infty,$$

where  $G''_k$  and  $G_k^{(j)}$  denote the second and  $j$ th derivatives of  $G_k$ , respectively.

- (D5) Let  $(\mathbf{Z}^{(k)}, \mathbf{Y}^{(k)})$  be the components of  $(\mathbf{Z}, \mathbf{Y})$  that appear in the regression of  $T_k$  ( $k = 1, \dots, K_1$ ). For any vectors  $\mathbf{h}_1$ ,  $\mathbf{h}_{2k}$ , and  $\mathbf{h}_{3k}$  of appropriate dimensions, if

$$\begin{aligned} & \frac{\partial}{\partial (\boldsymbol{\psi}, \mathbf{v})} f_{Y, \eta_1}(\mathbf{Y}, \boldsymbol{\eta}_1 \mid \mathbf{Z}; \boldsymbol{\psi}, \mathbf{v})^\top \mathbf{h}_1 \\ & - \sum_{k=1}^{K_1} (\mathbf{Z}^{(k)\top} \mathbf{h}_{2k} + \mathbf{Y}^{(k)\top} \mathbf{h}_{3k}) \frac{\partial}{\partial \eta_{1k}} f_{Y, \eta_1}(\mathbf{Y}, \boldsymbol{\eta}_1 \mid \mathbf{Z}; \boldsymbol{\psi}, \mathbf{v}) \end{aligned}$$

is equal to 0 for all  $\mathbf{Z}, \mathbf{Y}$ , and  $\boldsymbol{\eta}_1$ , then  $\mathbf{h}_1 = \mathbf{0}$ ,  $\mathbf{h}_{2k} = \mathbf{0}$ , and  $\mathbf{h}_{3k} = \mathbf{0}$  for  $k = 1, \dots, K_1$ , where  $f_{Y, \eta_1}$  is defined in condition (C4).

*Remark 5.* Conditions (D1)–(D4) are similar to the conditions of Zeng and Lin (2010) for joint modeling of longitudinal and survival data. Extra conditions are imposed on the transformations and the distributions of  $\mathbf{Y}$  and  $\boldsymbol{\eta}$  to accommodate the presence of unbounded covariate  $\mathbf{Y}$  in the survival model. Condition (D5) is for the invertibility of the information operator. If  $\boldsymbol{\alpha}_k = \mathbf{0}$  and  $\boldsymbol{\beta}_k = \mathbf{0}$  for  $k = 1, \dots, K_1$ , then condition (D5) simply requires that the information matrix of the model for  $(\mathbf{Y}, \boldsymbol{\eta}_1)$  is invertible. This is parallel to condition (C4) for identifiability.

*Remark 6.* The conditions for identifiability and the invertibility of the information operator (C1)–(C5), and (D5) differ significantly from the corresponding conditions (C5) and (C7) of Zeng and Lin (2010). The latter are stated under very general settings, but they are hard to verify for specific models, especially under our SEM framework. By contrast, our conditions are easier to verify and have intuitive interpretations. For the model in [Example 1](#), condition (D5) simply requires that the model of  $(Y_1, Y_2, Y_3, \eta_1)$  given  $Z$  has a nonzero score statistic, which clearly holds.

Let  $\mathcal{A}_0 = (\Lambda_{01}, \dots, \Lambda_{0K})$  and  $(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})$  be the NPMLE of  $(\boldsymbol{\theta}, \mathcal{A})$ . Also, let  $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^d, |\mathbf{v}| \leq 1\}$  and  $\mathcal{Q} = \{h(t) : \|h(t)\|_{V[0, \tau]} \leq 1\}$  with  $\|\cdot\|_{V[0, \tau]}$  being the total variation norm on  $[0, \tau]$ . We consider  $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \hat{\mathcal{A}} - \mathcal{A}_0)$  as a random element

in  $l^\infty(\mathcal{V} \times \mathcal{Q}^K)$  with

$$\begin{aligned} & (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \hat{\mathbf{A}} - \mathbf{A}_0)(\mathbf{v}, h_1, \dots, h_K) \\ &= (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbf{v} + \sum_{k=1}^K \int_0^\tau h_k(s) d(\hat{\Lambda}_k - \Lambda_{0k})(s). \end{aligned}$$

We have the following results.

*Theorem 2.* Under conditions (C1)–(C5) and (D1)–(D5),

1.  $|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| + \sum_{k=1}^K \sup_{t \in [0, \tau]} |\hat{\Lambda}_k(t) - \Lambda_{0k}(t)| \rightarrow_{a.s.} 0$ ; and
2.  $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \hat{\mathbf{A}} - \mathbf{A}_0) \rightarrow_d \mathcal{G}$  in  $l^\infty(\mathcal{V} \times \mathcal{Q}^K)$ , where  $\mathcal{G}$  is a continuous zero-mean Gaussian process. Furthermore, the limiting covariance matrix of  $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$  attains the semiparametric efficiency bound.

*Remark 7.* The proof of [Theorem 2](#) relies on the Donsker properties of certain classes of functions. It is more challenging to establish the Donsker results in our setting than in previous settings (e.g., Kosorok, Lee, and Fine 2004; Zeng and Lin 2010) because the likelihood function of the proposed model may contain the unbounded variable  $Y$ .

*Remark 8.* A key step in proving the asymptotic normality of the NPMLE is to show that the information operator is invertible. The result is given by Lemma 2 in the [Appendix](#), which states that condition (D5), together with conditions (C1)–(C3), and (C5), implies that the information operator of the model is invertible. With this result, we can verify the invertibility of the information operator of the semiparametric model by inspecting the parametric part of the model that contains the observed and latent variables. For the frailty models in Kosorok, Lee, and Fine (2004), verification of the invertibility of the information operator involves inspection of the local behavior of the score around  $T = 0$ . However, that approach is limited to frailty distributions that are indexed by a one-dimensional parameter and is not directly applicable to cases with more complex latent variable distributions such as those in our setting.

## 5. Simulation Studies

We considered a model with covariates  $\mathbf{Z} = (Z_1, Z_2)^\top$ , two latent variables  $(\eta_1, \eta_2)$ , observed continuous variables

$(Y_1, \dots, Y_5)$ , observed binary variables  $(Y_6, Y_7)$ , and a survival time  $T$ . Their distributions are given by

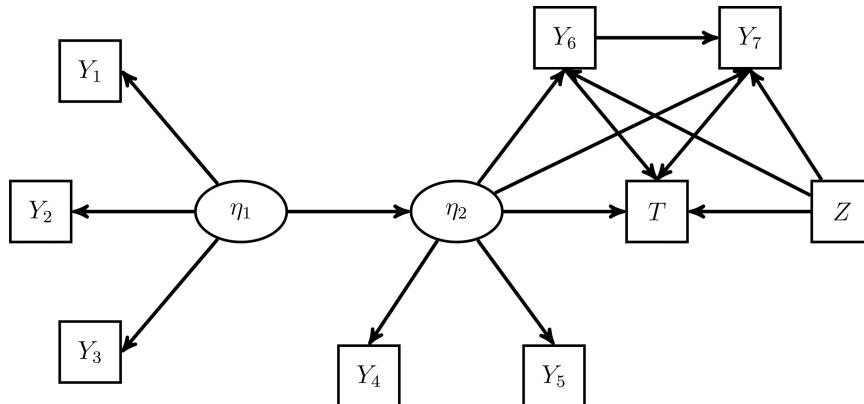
$$\begin{aligned} & \Lambda_T(t | \mathbf{Z}, Y_6, Y_7, \eta_2) \\ &= G\{\Lambda_0(t) \exp(\mathbf{X}_T^\top \boldsymbol{\beta}_T + \phi_T \eta_2)\}, \quad \mathbf{X}_T = (Z_1, Z_2, Y_6, Y_7)^\top, \\ & \text{logit}\{P(Y_6 = 1 | \mathbf{Z}, \eta_2)\} \\ &= \mathbf{X}_{Y_6}^\top \boldsymbol{\beta}_{Y_6} + \phi_{Y_6} \eta_2, \quad \mathbf{X}_{Y_6} = (1, Z_1, Z_2)^\top, \\ & \text{logit}\{P(Y_7 = 1 | \mathbf{Z}, Y_6, \eta_2)\} \\ &= \mathbf{X}_{Y_7}^\top \boldsymbol{\beta}_{Y_7} + \phi_{Y_7} \eta_2, \quad \mathbf{X}_{Y_7} = (1, Z_1, Z_2, Y_6)^\top, \\ & Y_j | \eta_1 \sim N(\beta_{Y_j} + \phi_{Y_j} \eta_1, \sigma_{Y_j}^2), \quad j = 1, 2, 3, \\ & Y_j | \eta_2 \sim N(\beta_{Y_j} + \phi_{Y_j} \eta_2, \sigma_{Y_j}^2), \quad j = 4, 5, \\ & \eta_2 | \eta_1 \sim N(\beta_{\eta_2} \eta_1, \sigma_{\eta_2}^2), \\ & \eta_1 \sim N(0, \sigma_{\eta_1}^2). \end{aligned}$$

The parameters  $\phi_{Y_1}$  and  $\phi_{Y_4}$  are fixed to be one. The model is depicted in [Figure 3](#).

We set  $Z_1$  and  $Z_2$  to independent standard normal and Bernoulli(0.5), respectively, and  $\Lambda_0(t) = t^2$ . We considered the class of logarithmic transformations  $G(x) = r^{-1} \log(1 + rx)$  with  $r = 0$  or 1, which correspond to the proportional hazards and proportional odds models, respectively. We generated the censoring times from  $\exp(c)$ , where  $c$  was chosen to yield approximately 30% censored observations. We set  $(Y_1, \dots, Y_5)$  to be missing completely at random for 30% of the subjects. We set the sample size to 400 and set the number of abscissa points to 20 for each Gauss-Hermite quadrature. We simulated 5000 datasets for each setting. The results are summarized in [Table 1](#).

The estimators of all parameters are virtually unbiased for both the proportional hazards and proportional odds models. The standard error estimators accurately reflect the true variations, and the coverage probabilities of the confidence intervals are close to the nominal level. Standard error estimators for the parameters in the survival model are larger under the proportional odds model than under the proportional hazards model. As a result, the standard error estimators for the parameters associated with  $\eta_2$  are larger. The standard error estimators for the remaining parameters are very similar between the two models.

We also evaluated Mplus (Muthén and Muthén 1998–2015) under the proportional hazards model, and the results are presented in Table S.1 of the supplementary materials. The results



**Figure 3.** Model used in simulation studies. The SEM consists of two latent variables, an observed covariate, seven binary or normal manifest variables, and a survival time that regresses on the latent variable, some manifest variables, and the observed covariates.

**Table 1.** Simulation results for the SEM with two latent variables.

Dep	Ind	Proportional hazards model					Proportional odds model				
		Param	Bias	SE	SEE	CP	Param	Bias	SE	SEE	CP
T	Z <sub>1</sub>	0.100	0.003	0.075	0.073	0.947	0.100	0.003	0.109	0.107	0.946
	Z <sub>2</sub>	-0.200	-0.005	0.180	0.181	0.953	-0.200	0.002	0.269	0.270	0.950
	Y <sub>6</sub>	0.100	0.000	0.145	0.142	0.947	0.100	0.001	0.212	0.208	0.944
	Y <sub>7</sub>	0.200	-0.001	0.165	0.163	0.949	0.200	0.002	0.238	0.236	0.950
	η <sub>2</sub>	0.500	0.012	0.167	0.166	0.950	0.500	0.014	0.228	0.223	0.948
	Λ <sub>0</sub> (t <sub>1</sub> )	0.202	0.004	0.050	0.049	0.949	0.212	0.006	0.074	0.072	0.947
	Λ <sub>0</sub> (t <sub>2</sub> )	0.518	0.010	0.120	0.118	0.949	0.608	0.020	0.206	0.202	0.953
	Λ <sub>0</sub> (t <sub>3</sub> )	1.103	0.032	0.256	0.249	0.950	1.588	0.076	0.554	0.538	0.950
Y <sub>1</sub>	Int	0.000	-0.001	0.073	0.073	0.947	0.000	-0.001	0.073	0.073	0.948
	Var	1.000	-0.010	0.131	0.128	0.960	1.000	-0.010	0.131	0.128	0.960
Y <sub>2</sub>	Int	0.000	0.001	0.074	0.073	0.945	0.000	0.001	0.074	0.073	0.945
	η <sub>1</sub>	1.000	0.021	0.210	0.203	0.949	1.000	0.022	0.210	0.204	0.950
Y <sub>3</sub>	Var	1.000	-0.013	0.129	0.128	0.960	1.000	-0.013	0.129	0.128	0.960
	Int	0.000	0.000	0.073	0.073	0.954	0.000	0.000	0.073	0.073	0.954
Y <sub>4</sub>	η <sub>1</sub>	1.000	0.024	0.211	0.204	0.950	1.000	0.024	0.212	0.204	0.952
	Var	1.000	-0.015	0.131	0.128	0.959	1.000	-0.015	0.131	0.128	0.960
Y <sub>5</sub>	Int	0.000	0.001	0.077	0.076	0.946	0.000	0.001	0.077	0.076	0.946
	Var	1.000	-0.025	0.194	0.185	0.956	1.000	-0.029	0.216	0.209	0.951
Y <sub>6</sub>	Int	0.000	0.001	0.077	0.076	0.950	0.000	0.001	0.077	0.076	0.949
	η <sub>2</sub>	1.000	0.048	0.303	0.279	0.942	1.000	0.060	0.348	0.323	0.945
Y <sub>7</sub>	Var	1.000	-0.030	0.198	0.188	0.957	1.000	-0.036	0.221	0.213	0.954
	Int	0.000	0.004	0.278	0.273	0.953	0.000	0.004	0.277	0.273	0.954
η <sub>1</sub>	Z <sub>1</sub>	-0.500	-0.006	0.113	0.113	0.953	-0.500	-0.006	0.113	0.113	0.953
	Z <sub>2</sub>	0.500	0.006	0.301	0.297	0.954	0.500	0.006	0.300	0.297	0.955
	η <sub>2</sub>	0.000	0.003	0.229	0.220	0.964	0.000	0.003	0.231	0.222	0.965
	Int	1.000	0.033	0.347	0.345	0.953	1.000	0.033	0.347	0.345	0.953
	Z <sub>1</sub>	1.000	0.024	0.151	0.148	0.951	1.000	0.024	0.151	0.148	0.951
	Z <sub>2</sub>	0.200	-0.005	0.344	0.342	0.955	0.200	-0.005	0.345	0.342	0.955
	Y <sub>6</sub>	-0.200	-0.009	0.265	0.264	0.952	-0.200	-0.009	0.265	0.264	0.953
	η <sub>2</sub>	0.000	0.003	0.265	0.253	0.960	0.000	0.002	0.268	0.255	0.962
η <sub>2</sub>	Var	0.500	0.006	0.136	0.134	0.956	0.500	0.006	0.136	0.134	0.955
	η <sub>1</sub>	0.500	0.009	0.153	0.147	0.939	0.500	0.009	0.157	0.151	0.940
	Var	0.500	0.009	0.177	0.169	0.960	0.500	0.013	0.196	0.190	0.964

NOTE: Each row corresponds to the regression parameter of the dependent variable “Dep” on the independent variable “Ind” or some other parameter in the model of “Dep.” “Int” and “Var” stand for the intercept and error variance, respectively. The parameters  $\Lambda_0(t_1)$ ,  $\Lambda_0(t_2)$ , and  $\Lambda_0(t_3)$  correspond to the cumulative baseline hazard function values at the 25%, 50%, and 75% quantiles of the survival time. The true value of a parameter is given under “Param.” “Bias” is the empirical bias; “SE” is the empirical standard error; “SEE” is the empirical mean of the standard error estimator; and “CP” is the empirical coverage probability of the 95% confidence interval.

for the Euclidean parameters are similar to those presented in Table 1. Mplus provides estimator for the baseline hazard function instead of the cumulative baseline hazard function. Its standard error estimator does not reflect the true variation, and the coverages of the confidence intervals are far below the nominal level.

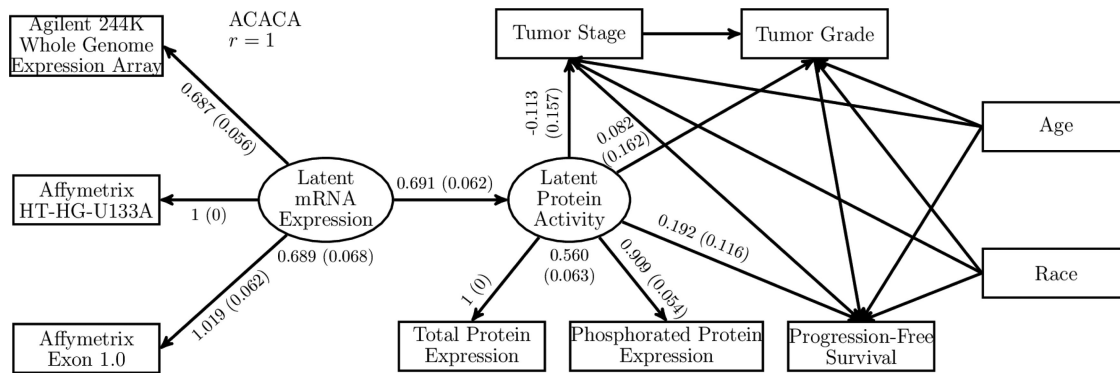
## 6. Real Data Analysis

We analyzed a dataset on patients with serous ovarian cancer from the TCGA project (The Cancer Genome Atlas Research Network 2011). Genomic variables include DNA copy number, SNP genotype, DNA methylation level, and levels of expression of mRNA, microRNA, total protein, and phosphorylated protein. Demographic and clinical variables include age at diagnosis, race, tumor stage, tumor grade, time to tumor progression, and time to death. There are a total of 586 patients. The median follow-up time was about 2.5 years, and roughly 30% of the patients were lost to follow-up before tumor progression or death. The data are available from <http://gdac.broadinstitute.org/>.

We focused on the integrative analysis of clinical outcomes and expression levels of mRNA, total protein, and phosphorylated protein. We considered mRNA expression as a latent variable that can only be observed with error through three

microarray platforms, namely Agilent 244K Whole Genome Expression Array, Affymetrix HT-HG-U133A, and Affymetrix Exon 1.0. We assumed that the effects of a gene on clinical outcomes are mediated through unobserved protein activity. The latent protein activity is modified by mRNA expression and is manifest through the observed protein expression measurements, which were obtained from the reverse-phase protein arrays platform. Figure 4 depicts the SEM fit for each gene. We assumed that the observed variables follow the distributions described in Section 5, with  $(Y_1, Y_2, Y_3)$  being the three microarray measurements,  $(Y_4, Y_5) = (\text{Total protein expression, Phosphorylated protein expression})$ ,  $(Y_6, Y_7) = (\text{Tumor stage, Tumor grade})$ ,  $(Z_1, Z_2) = (\text{Age, Race})$ , and  $T$  being progression-free survival time.

We dichotomized tumor stage into stage II/III versus stage IV and tumor grade into grade 2 versus grade 3/4. Race was dichotomized into white and nonwhite. We allowed mRNA expression and protein expression data to be missing for some subjects. We excluded patients with tumor stage I or grade 1, as those patients may have a disease that is biologically different from that of patients with tumors of other stages or grades. For each gene, we fit the class of transformation models with  $G(x) = r^{-1} \log(1 + rx)$  over a grid of  $r = (0, 0.1, \dots, 2)$ . We selected the model with the smallest AIC or, equivalently, the largest log-likelihood value.



**Figure 4.** Results from the SEM analysis of the gene ACACA. Analysis results are from 542 patients with ovarian cancer in the TCGA project. The numbers besides an arrow correspond to the point estimate and standard error estimate (in parentheses) of the regression parameter. The numbers below the latent variables correspond to the point estimate and standard error estimate (in parentheses) of the error variance.

We present the results for the gene ACACA. The sample size is 542. About 30% of the subjects do not have protein expression data, and over 10% of the subjects miss at least one mRNA expression measurement. The best-fitting model is obtained at  $r = 1$ , which corresponds to the proportional odds model. The point estimates and standard error estimates of the parameters associated with the latent variables are shown in Figure 4. The remaining results are shown in Table S.2 of the Supplementary Materials. The latent variables have strong positive association with the measurement platforms. As expected, latent protein activity and latent mRNA expression are highly correlated. Latent protein activity is positively associated with progression-free survival time, with a  $p$ -value of 0.100. Specifically, higher latent protein activity is associated with shorter progression-free survival time, which agrees with the findings of the literature (Menendez and Lupu 2007). The association of ACACA with tumor stage or tumor grade is weak.

The results for the parameters in the nonsurvival models are similar between  $r = 0$  and 1. The parameters in the survival model have different interpretations between  $r = 0$  and 1. With  $r = 0$ , a unit increase in latent protein activity would have a multiplicative effect of  $\exp(0.068)$  on the hazard function. With  $r = 1$ , a unit increase in the latent protein activity would have a multiplicative effect of  $\exp(-0.192)$  on the survival odds. For this dataset, the proportional odds model provides much stronger evidence for the effect of protein activity on progression-free survival than the proportional hazards model.

For the Cox proportional hazards model, we also present the results from Mplus in Table S.2. The results from NPMLE and Mplus are similar for most parameters. There are considerable differences between the cumulative baseline hazard function estimates. The standard error estimates for the cumulative baseline hazard function are not available from Mplus.

For comparisons, we also fit a proportional odds model without latent variables for progression-free survival on the covariates and the two protein expression variables, where the subjects with missing protein expression data were discarded. The  $p$ -value of the Wald test for the joint effect of protein expression is 0.157. With  $r = 0$ , the Wald test  $p$ -value is 0.578. Therefore, analyses based on standard models fail to conclude a strong association between the protein expression and progression-free survival. The power of the proposed SEM

framework stems from the appropriate handling of missing data, the dimension reduction of the observed covariates, and the flexibility of the survival model.

## 7. Discussion

In this article, we consider semiparametric SEM for potentially right-censored survival time data. We prove the consistency, asymptotic normality, and semiparametric efficiency of the NPMLE. We propose new rules for establishing model identifiability and invertibility of the information operator. We construct an EM algorithm to compute the NPMLE and introduce occasional Newton-Raphson steps to accelerate the convergence.

One contribution of Theorem 1 is that it reduces a semi-parametric identifiability problem to a parametric one; it shows that the inclusion of the semiparametric component does not make the model less identifiable but, in some sense, makes the model more easily identifiable. With that being said, the result hinges on correct specification of the model and does not guarantee empirical identifiability in a finite sample. Therefore, care should be taken when fitting a model that is nearly nonidentifiable. Another main result of ours is given by Lemma 1. This lemma is applicable to a wide range of latent variable models and allows one to deduce the identifiability of a model by inspecting just part of it.

Invertibility of the information operator has received much less attention in the literature than model identifiability. In this article, we prove a general result for invertibility of the information operator. It is evident from the proof that the invertibility of the information operator can be established using techniques similar to those used to establish model identifiability. Specifically, the key to the proof of the identifiability of the mixture Cox model is that with the presence of a covariate that is independent of the latent variable, the contributions to the likelihood from the latent variable and the baseline hazard function can be separated by considering different values of the covariate. (In a normal mixture model, however, we lack such identifiability results precisely because the random effect and error term are combined linearly and their distributions cannot be distinguished.) As a result, if two sets of parameters give rise to the same marginal survival function, then they must do so by giving rise to the same random-effect distribution. Based on



the proportional hazards structure, we prove a parallel result for the invertibility of the information operator: the existence of a submodel with zero score implies that the random-effect distribution has zero score along that submodel as well. Therefore, to ensure the invertibility of the information operator of the mixture Cox model, one only has to ensure that the information matrix of the random-effect distribution is invertible.

Our work can be extended in several directions. First, one may be interested in expanding the model by inclusion of more latent and observed variables. As the number of variables increases, the number of parameters to be estimated increases as well. Then, it may be desirable to perform variable selection. Because a single variable may be associated with multiple parameters, one may prefer not to treat parameters as the basic unit of selection, as in traditional lasso methods (Tibshirani 1996). Instead, methods like group lasso (Yuan and Lin 2006) that penalize parameters associated with a variable as a group may be considered.

In our model, the distribution of the manifest variable  $\mathbf{Y}$  is fully parametric. One can allow a nonparametric transformation on  $\mathbf{Y}$ . A major challenge arises in extending the asymptotic results to unbounded nonparametric transformation, as the estimator of the transformation function can be unbounded (Zeng and Lin 2010).

Finally, it would be of interest to consider interval-censored data. Interval censoring results in a different likelihood function, which makes the computation of the NPMLE and the derivation of its asymptotic properties challenging, even for univariate survival time data. The asymptotic theory for interval-censored data is only available in a few simple cases; see Huang and Wellner (1997) for a review.

## Appendix: Technical Details

We present the following conditions, which are clearly implied by conditions (C3)–(C5):

(C3') For  $k = 1, \dots, K_1$ ,  $\boldsymbol{\vartheta}_k$  is a nonzero vector.

(C4') Consider two sets of parameters  $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \boldsymbol{\psi}, \boldsymbol{\nu})$  and  $(\tilde{\boldsymbol{\alpha}}_k, \tilde{\boldsymbol{\beta}}_k, \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}})$  for  $k = 1, \dots, K_1$ , and let  $\tilde{\boldsymbol{\eta}}_1 = (\tilde{\eta}_{11}, \dots, \tilde{\eta}_{1K_1})$ , where  $\tilde{\eta}_{1k} = \mathbf{Y}^\top(\boldsymbol{\alpha}_k - \tilde{\boldsymbol{\alpha}}_k) + \mathbf{Z}^\top(\boldsymbol{\beta}_k - \tilde{\boldsymbol{\beta}}_k) + \eta_{1k}$ . Let  $f_{Y, \boldsymbol{\eta}_1}$  be the density of  $(\mathbf{Y}, \boldsymbol{\eta}_1)$  given  $\mathbf{Z}$ . Then,  $f_{Y, \boldsymbol{\eta}_1}(\mathbf{Y}, \boldsymbol{\eta}_1 | \mathbf{Z}; \boldsymbol{\psi}, \boldsymbol{\nu}) = f_{Y, \boldsymbol{\eta}_1}(\mathbf{Y}, \tilde{\boldsymbol{\eta}}_1 | \mathbf{Z}; \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}})$  for all  $\mathbf{Z}$ ,  $\mathbf{Y}$ , and  $\boldsymbol{\eta}_1$  implies that  $\boldsymbol{\alpha}_k = \tilde{\boldsymbol{\alpha}}_k$ ,  $\boldsymbol{\beta}_k = \tilde{\boldsymbol{\beta}}_k$ ,  $\boldsymbol{\psi} = \tilde{\boldsymbol{\psi}}$ , and  $\boldsymbol{\nu} = \tilde{\boldsymbol{\nu}}$ .

(C5') For  $k = K_1 + 1, \dots, K$ , let  $(\mathbf{Y}^{(k)}, \boldsymbol{\eta}_1^{(k)}, \boldsymbol{\eta}_2^{(k)})$  be the components of  $(\mathbf{Y}, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$  that appear in the regression of  $T_k$ , and let  $(\mathbf{Y}^{-(k)}, \boldsymbol{\eta}_1^{-(k)}, \boldsymbol{\eta}_2^{-(k)})$  be the remaining components. If  $\boldsymbol{\eta}_2^{(k)}$  is nonempty, then  $(\mathbf{Y}^{-(k)}, \boldsymbol{\eta}_1^{-(k)})$  is nonempty, and  $\boldsymbol{\eta}_2^{(k)}$  is complete sufficient in  $\{F_{\boldsymbol{\eta}_2^{(k)} | \mathbf{Y}, \boldsymbol{\eta}_1}(\cdot | \mathbf{Z}, \mathbf{Y}, \boldsymbol{\eta}_1) : \mathbf{Z} = \mathbf{z}_0, \mathbf{Y}^{(k)} = \mathbf{y}_0, \boldsymbol{\eta}_1^{(k)} = \boldsymbol{\eta}_{10}\}$  for any fixed  $\mathbf{z}_0, \mathbf{y}_0$ , and  $\boldsymbol{\eta}_{10}$ , where  $F_{\boldsymbol{\eta}_2^{(k)} | \mathbf{Y}, \boldsymbol{\eta}_1}$  is the distribution function of  $\boldsymbol{\eta}_2^{(k)}$  given  $(\mathbf{Z}, \mathbf{Y}, \boldsymbol{\eta}_1)$  with  $(\mathbf{Y}^{-(k)}, \boldsymbol{\eta}_1^{-(k)})$  treated as a parameter vector.

We prove Theorem 1 under the generalized conditions (C1), (C2), and (C3')–(C5'). The proof makes use of two lemmas given at the end of this Appendix. We first provide an overview of the proof. For any two sets of parameters  $(\boldsymbol{\vartheta}_k, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \boldsymbol{\phi}_k, \Lambda_k, \boldsymbol{\psi}, \boldsymbol{\nu})$  and  $(\tilde{\boldsymbol{\vartheta}}_k, \tilde{\boldsymbol{\alpha}}_k, \tilde{\boldsymbol{\beta}}_k, \tilde{\boldsymbol{\phi}}_k, \tilde{\Lambda}_k, \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}})$ , assume that the likelihood values at the two sets of parameters are identical almost surely. By definition, the

model is identifiable if the equality of the likelihood values implies the equality of the two sets of parameters. We derive the equality of the two sets of parameters in the following steps:

1. By conditions (C1), (C2), and (C3') and the identifiability of the mixture Cox model (Kortram et al. 1995),  $\boldsymbol{\vartheta}_k = \tilde{\boldsymbol{\vartheta}}_k$  and  $\Lambda_k = \tilde{\Lambda}_k$  for  $k = 1, \dots, K_1$ .
2. With some algebraic manipulation, the likelihood function can be expressed in the form of the Laplace transform of the distribution of a function of  $(\mathbf{Y}, \boldsymbol{\eta}_1)$ . The uniqueness of the Laplace transform, together with condition (C4'), implies that  $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \boldsymbol{\psi}, \boldsymbol{\nu}) = (\tilde{\boldsymbol{\alpha}}_k, \tilde{\boldsymbol{\beta}}_k, \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}})$  for  $k = 1, \dots, K_1$ .
3. By the uniqueness of the Laplace transform and the complete sufficiency of  $\boldsymbol{\eta}_2$  imposed by condition (C5'), the equality of the likelihood functions of  $(T_{K_1+1}, \dots, T_K, \mathbf{Y})$  implies the equality of the likelihood functions of  $(T_{K_1+1}, \dots, T_K, \mathbf{Y}, \boldsymbol{\eta})$ . By the identifiability of the Cox model, we conclude that  $(\boldsymbol{\vartheta}_k, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \boldsymbol{\phi}_k) = (\tilde{\boldsymbol{\vartheta}}_k, \tilde{\boldsymbol{\alpha}}_k, \tilde{\boldsymbol{\beta}}_k, \tilde{\boldsymbol{\phi}}_k)$  for  $k = K_1 + 1, \dots, K$ .

*Proof of Theorem 1.* The likelihood is given in (4). Here, we consider a single observation and drop the subscript  $i$ . Using the arguments in Section 10.1 of Zeng and Lin (2010), we can set each survival time to be right censored at any time point within  $[0, \tau]$  when establishing identifiability. Consider two sets of parameters  $(\boldsymbol{\vartheta}_k, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \boldsymbol{\phi}_k, \Lambda_k, \boldsymbol{\psi}, \boldsymbol{\nu})$  and  $(\tilde{\boldsymbol{\vartheta}}_k, \tilde{\boldsymbol{\alpha}}_k, \tilde{\boldsymbol{\beta}}_k, \tilde{\boldsymbol{\phi}}_k, \tilde{\Lambda}_k, \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}})$  such that the likelihood values for an observation with the  $K$  survival times being right censored are equal almost surely, that is,

$$\begin{aligned} & \int \prod_{k=1}^K \left[ \int \exp \left\{ -\Lambda_k(t_k) s_k e^{\mathbf{W}^\top \boldsymbol{\vartheta}_k + \mathbf{Z}^\top \boldsymbol{\beta}_k + \mathbf{Y}^\top \boldsymbol{\alpha}_k + \boldsymbol{\eta}^\top \boldsymbol{\phi}_k} \right\} \right. \\ & \quad \times g_k(s_k) dm_k(s_k) \left. \right] f_{Y, \boldsymbol{\eta}}(\mathbf{Y}, \boldsymbol{\eta} | \mathbf{Z}; \boldsymbol{\psi}, \boldsymbol{\nu}) d\boldsymbol{\eta} \\ & = \int \prod_{k=1}^K \left[ \int \exp \left\{ -\tilde{\Lambda}_k(t_k) s_k e^{\mathbf{W}^\top \tilde{\boldsymbol{\vartheta}}_k + \mathbf{Z}^\top \tilde{\boldsymbol{\beta}}_k + \mathbf{Y}^\top \tilde{\boldsymbol{\alpha}}_k + \boldsymbol{\eta}^\top \tilde{\boldsymbol{\phi}}_k} \right\} \right. \\ & \quad \times g_k(s_k) dm_k(s_k) \left. \right] f_{Y, \boldsymbol{\eta}}(\mathbf{Y}, \boldsymbol{\eta} | \mathbf{Z}; \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}}) d\boldsymbol{\eta} \end{aligned} \quad (\text{A.1})$$

for all  $t_1, \dots, t_K \in [0, \tau]$ ,  $\mathbf{W}$ ,  $\mathbf{Z}$ , and  $\mathbf{Y}$ , where  $f_{Y, \boldsymbol{\eta}}$  is the density of  $(\mathbf{Y}, \boldsymbol{\eta})$  given  $\mathbf{Z}$ . If  $m_k$  is a point mass at one, then  $s_k$  is fixed at one,  $g_k = 1$ , and the integration with respect to  $m_k(s_k)$  can be omitted. For simplicity of description, assume that  $m_k$  is the Lebesgue measure. Note that

$$\begin{aligned} \int s g_k(s) ds & = - \lim_{t \rightarrow 0^+} \frac{d}{dt} \int_0^\infty e^{-ts} g_k(s) ds \\ & = - \lim_{t \rightarrow 0^+} \frac{d}{dt} \exp\{-G_k(t)\} = G'_k(0) < \infty. \end{aligned}$$

Thus, a transformation model can be written as a random-effect proportional hazards model with known distributions  $(g_1, \dots, g_K)$  for random effects  $(s_1, \dots, s_K)$  with finite means.

First, we show that the baseline hazard functions of the first  $K_1$  survival times are identifiable. For each  $k = 1, \dots, K_1$ , set  $t_l \rightarrow 0$  for  $l \neq k$  on both sides of (A.1). On each side of the resulting equation, integration with respect to  $\mathbf{Y}$  results in the likelihood of a mixture Cox model with  $s_k e^{\mathbf{Y}^\top \boldsymbol{\alpha}_k + \boldsymbol{\eta}_k}$  or  $s_k e^{\mathbf{Y}^\top \tilde{\boldsymbol{\alpha}}_k + \boldsymbol{\eta}_k}$  as a latent variable. Let  $E(\cdot | \mathbf{Z})$  and  $\tilde{E}(\cdot | \mathbf{Z})$  be the expectations under  $f_{Y, \boldsymbol{\eta}}(\cdot | \mathbf{Z}; \boldsymbol{\psi}, \boldsymbol{\nu})$  and  $f_{Y, \boldsymbol{\eta}}(\cdot | \mathbf{Z}; \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}})$ , respectively. Theorem

3 of Kortram et al. (1995) implies that  $E(s_k e^{Y^T \alpha_k + \eta_{1k}} | \mathbf{Z} = \mathbf{0}) \Lambda_k = \tilde{E}(s_k e^{Y^T \tilde{\alpha}_k + \eta_{1k}} | \mathbf{Z} = \mathbf{0}) \tilde{\Lambda}_k$  on  $[0, \tau]$ ,  $\vartheta_k = \tilde{\vartheta}_k$ , and the distribution of  $E(s_k e^{Y^T \alpha_k + \eta_{1k}} | \mathbf{Z} = \mathbf{0})^{-1} s_k e^{Y^T \alpha_k + \eta_{1k}}$  under  $f_{Y, \eta}(\cdot | \mathbf{Z}; \boldsymbol{\psi}, \boldsymbol{\nu})$  is equal to that of  $\tilde{E}(s_k e^{Y^T \tilde{\alpha}_k + \eta_{1k}} | \mathbf{Z} = \mathbf{0})^{-1} s_k e^{Y^T \tilde{\alpha}_k + \eta_{1k}}$  under  $f_{Y, \eta}(\cdot | \mathbf{Z}; \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}})$ . Because  $E(\mathbf{Y}^T \boldsymbol{\alpha}_k + \eta_{1k} | \mathbf{Z} = \mathbf{0}) = \tilde{E}(\mathbf{Y}^T \tilde{\boldsymbol{\alpha}}_k + \eta_{1k} | \mathbf{Z} = \mathbf{0}) = 0$  by condition (C2), we see that  $\Lambda_k = \tilde{\Lambda}_k$  on  $[0, \tau]$ .

Second, we show that the likelihood function takes the form of a Laplace transform and use the uniqueness of the Laplace transform to prove the identifiability of  $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \boldsymbol{\psi}, \boldsymbol{\nu})$  ( $k = 1, \dots, K_1$ ). Setting  $t_k \rightarrow 0$  for  $k = K_1 + 1, \dots, K$  and  $\mathbf{W} = \mathbf{0}$  on both sides of (A.1), we have

$$\begin{aligned} & \int \prod_{k=1}^{K_1} \left[ \int \exp \left\{ -\Lambda_k(t_k) s_k e^{Z^T \boldsymbol{\beta}_k + Y^T \boldsymbol{\alpha}_k + \eta_{1k}} \right\} g_k(s_k) ds_k \right] \\ & \quad \times f_{Y, \eta}(\mathbf{Y}, \boldsymbol{\eta} | \mathbf{Z}; \boldsymbol{\psi}, \boldsymbol{\nu}) d\boldsymbol{\eta} \\ & = \int \prod_{k=1}^{K_1} \left[ \int \exp \left\{ -\Lambda_k(t_k) s_k e^{Z^T \tilde{\boldsymbol{\beta}}_k + Y^T \tilde{\boldsymbol{\alpha}}_k + \eta_{1k}} \right\} g_k(s_k) ds_k \right] \\ & \quad \times f_{Y, \eta}(\mathbf{Y}, \boldsymbol{\eta} | \mathbf{Z}; \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}}) d\boldsymbol{\eta}. \end{aligned} \quad (\text{A.2})$$

Let  $\mathbf{U} = (U_1, \dots, U_{K_1})$ ,  $U_k = s_k e^{\eta_{1k}}$ , and  $f_{U|Y}$  be the density function of  $\mathbf{U}$  given  $\mathbf{Z}$  and  $\mathbf{Y}$ . By the uniqueness of the Laplace transform, for any continuous functions  $f$  and  $\tilde{f}$ , any open set  $\mathcal{S}$ , and any positive real numbers  $c$  and  $\tilde{c}$ ,

$$\int_0^\infty e^{-cst} f(t) dt = \int_0^\infty e^{-\tilde{c}st} \tilde{f}(t) dt \quad \forall s \in \mathcal{S}$$

implies that  $f(t) = (c/\tilde{c}) \tilde{f}(ct/\tilde{c})$  for all  $t > 0$ . Therefore, the equality of (A.2) for all  $t_1, \dots, t_{K_1}$ ,  $\mathbf{Z}$ , and  $\mathbf{Y}$  implies that

$$f_{U|Y}(\mathbf{U} | \mathbf{Z}, \mathbf{Y}; \boldsymbol{\psi}, \boldsymbol{\nu}) = e^{\sum_{k=1}^{K_1} Z^T (\boldsymbol{\beta}_k - \tilde{\boldsymbol{\beta}}_k) + Y^T (\boldsymbol{\alpha}_k - \tilde{\boldsymbol{\alpha}}_k)} f_{U|Y}(\tilde{\mathbf{U}} | \mathbf{Z}, \mathbf{Y}; \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}}) \quad (\text{A.3})$$

for all  $\mathbf{U}$ ,  $\mathbf{Z}$ , and  $\mathbf{Y}$ , where  $\tilde{\mathbf{U}} = (\tilde{U}_1, \dots, \tilde{U}_{K_1})$ , and  $\tilde{U}_k = e^{Z^T (\boldsymbol{\beta}_k - \tilde{\boldsymbol{\beta}}_k) + Y^T (\boldsymbol{\alpha}_k - \tilde{\boldsymbol{\alpha}}_k)} U_k$ . Let  $f_{\eta_1|Y}$  be the density of  $\eta_1$  given  $\mathbf{Z}$  and  $\mathbf{Y}$ . By the definition of  $\mathbf{U}$ ,

$$\begin{aligned} & f_{U|Y}(\mathbf{U} | \mathbf{Z}, \mathbf{Y}; \boldsymbol{\psi}, \boldsymbol{\nu}) \\ & = \int_{s_k > 0} f_{\eta_1|Y}(\log U_1 - \log s_1, \dots, \log U_{K_1} - \log s_{K_1} | \mathbf{Z}, \mathbf{Y}; \boldsymbol{\psi}, \boldsymbol{\nu}) \\ & \quad \times \prod_{k=1}^{K_1} U_k^{-1} g_k(s_k) d(s_1, \dots, s_{K_1}) \\ & = \int_{\mathbb{R}^{K_1}} f_{\eta_1|Y}(\log U_1 - v_1, \dots, \log U_{K_1} - v_{K_1} | \mathbf{Z}, \mathbf{Y}; \boldsymbol{\psi}, \boldsymbol{\nu}) \\ & \quad \times \prod_{k=1}^{K_1} U_k^{-1} \bar{g}_k(v_k) e^{v_k} d(v_1, \dots, v_{K_1}), \end{aligned}$$

where  $\bar{g}_k(v) = g_k(e^v)$ . Thus, (A.3) implies that

$$\begin{aligned} & \int_{\mathbb{R}^{K_1}} \left\{ f_{\eta_1|Y}(\log U_1 - v_1, \dots, \log U_{K_1} - v_{K_1} | \mathbf{Z}, \mathbf{Y}; \boldsymbol{\psi}, \boldsymbol{\nu}) \right. \\ & \quad \left. - f_{\eta_1|Y}(\log \tilde{U}_1 - v_1, \dots, \log \tilde{U}_{K_1} - v_{K_1} | \mathbf{Z}, \mathbf{Y}; \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}}) \right\} \\ & \quad \times \prod_{k=1}^{K_1} \bar{g}_k(v_k) e^{v_k} d(v_1, \dots, v_{K_1}) = 0. \end{aligned} \quad (\text{A.4})$$

Consider two arbitrary continuous functions  $f, g: \mathbb{R} \rightarrow \mathbb{R}$ . Note that

$$\int_{-\infty}^{\infty} e^{-st} (f * g)(t) dt = \int_{-\infty}^{\infty} e^{-st} f(t) dt \int_{-\infty}^{\infty} e^{-st} g(t) dt$$

for any  $s$  such that the integrals are defined, where  $(f * g)(t) \equiv \int_{-\infty}^{\infty} f(t-s)g(s) ds$  is the convolution of  $f$  and  $g$ . Therefore,  $(f * g)(\cdot) = 0$  implies that

$$\int_{-\infty}^{\infty} e^{-st} f(t) dt \int_{-\infty}^{\infty} e^{-st} g(t) dt = 0,$$

which, if  $g$  is positive, implies that  $f(\cdot) = 0$  by the uniqueness of the bilateral Laplace transform (Chareka 2007). Because  $\bar{g}_k(\cdot) e^{(\cdot)}$  is positive, (A.4) implies that  $f_{\eta_1|Y}(\eta_1 | \mathbf{Z}, \mathbf{Y}; \boldsymbol{\psi}, \boldsymbol{\nu}) = f_{\eta_1|Y}(\tilde{\eta}_1 | \mathbf{Z}, \mathbf{Y}; \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}})$ , where  $\tilde{\eta}_1$  is defined in condition (C4'). By condition (C4'),  $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \boldsymbol{\psi}, \boldsymbol{\nu}) = (\tilde{\boldsymbol{\alpha}}_k, \tilde{\boldsymbol{\beta}}_k, \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{\nu}})$  for  $k = 1, \dots, K_1$ .

It remains to identify the parameters associated with  $(T_{K_1+1}, \dots, T_K)$ . By the uniqueness of the Laplace transform, (A.1) implies that

$$\begin{aligned} & \int \prod_{k=K_1+1}^K \left\{ \int e^{-\Lambda_k(t_k) s_k e^{W^T \boldsymbol{\theta}_k + Z^T \boldsymbol{\beta}_k + Y^T \boldsymbol{\alpha}_k + \eta^T \boldsymbol{\phi}_k}} g_k(s_k) ds_k \right\} \\ & \quad \times f_{Y, \eta}(\mathbf{Y}, \boldsymbol{\eta} | \mathbf{Z}; \boldsymbol{\psi}, \boldsymbol{\nu}) d\boldsymbol{\eta}_2 \\ & = \int \prod_{k=K_1+1}^K \left\{ \int e^{-\tilde{\Lambda}_k(t_k) s_k e^{W^T \tilde{\boldsymbol{\theta}}_k + Z^T \tilde{\boldsymbol{\beta}}_k + Y^T \tilde{\boldsymbol{\alpha}}_k + \eta^T \tilde{\boldsymbol{\phi}}_k}} g_k(s_k) ds_k \right\} \\ & \quad \times f_{Y, \eta}(\mathbf{Y}, \boldsymbol{\eta} | \mathbf{Z}; \boldsymbol{\psi}, \boldsymbol{\nu}) d\boldsymbol{\eta}_2 \end{aligned}$$

for all  $t_{K_1+1}, \dots, t_K$ ,  $\mathbf{W}$ ,  $\mathbf{Z}$ ,  $\mathbf{Y}$ , and  $\boldsymbol{\eta}_1$ , that is,  $\boldsymbol{\eta}_1$  can be treated as observed for identifying the remaining parameters. Under condition (C5'), we can use the arguments in the proof of Lemma 1 to show that the integrands in the above equality are equal at each value of  $\boldsymbol{\eta}_2$ . We conclude that  $(\boldsymbol{\theta}_k, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \boldsymbol{\phi}_k, \Lambda_k) = (\tilde{\boldsymbol{\theta}}_k, \tilde{\boldsymbol{\alpha}}_k, \tilde{\boldsymbol{\beta}}_k, \tilde{\boldsymbol{\phi}}_k, \tilde{\Lambda}_k)$  for  $k = K_1 + 1, \dots, K$ .  $\square$

We provide an overview for the proof of Theorem 2. The consistency of the NPMLE is proved in the following steps:

1. By conditions (D2)–(D4), the NPMLE exists, that is,  $\hat{\Lambda}_k(\tau) < \infty$ .
2. By conditions (D3) and (D4),  $\hat{\Lambda}_k(\tau)$  is uniformly bounded. Helly's selection theorem then implies that every subsequence of  $\hat{\Lambda}_k$  has a further converging subsequence.
3. By the Glivenko-Cantelli properties of the log-likelihood and related functions given by Lemma S2 in the Supplementary Materials, the identifiability of the model, and the non-negativity of the Kullback-Leibler divergence, we conclude the consistency of the NPMLE.

The asymptotic normality of the NPMLE follows mainly from the arguments of van der Vaart (1998, pp. 419–424). Donsker properties of the score and related functions are given by Lemma S2 in the supplementary materials, and the invertibility of the information operator is given by Lemma 2.

*Proof of Theorem 2.* We use  $\mathbf{Z}$  to denote both  $\mathbf{W}$  and  $\mathbf{Z}$  with  $\boldsymbol{\beta}_k$  ( $k = 1, \dots, K$ ) being the corresponding vector of regression parameters.

Let

$$\begin{aligned} \Psi(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A}) &= \prod_{k=1}^K \int \left[ e^{\mathbf{Z}_i^T \boldsymbol{\beta}_k + \mathbf{Y}_i^T \boldsymbol{\alpha}_k + \boldsymbol{\eta}^T \boldsymbol{\phi}_k} G'_k \left\{ e^{\mathbf{Z}_i^T \boldsymbol{\beta}_k + \mathbf{Y}_i^T \boldsymbol{\alpha}_k + \boldsymbol{\eta}^T \boldsymbol{\phi}_k} \Lambda_k(\tilde{T}_{ki}) \right\} \right]^{\Delta_{ki}} \\ &\quad \times \exp \left[ -G_k \left\{ e^{\mathbf{Z}_i^T \boldsymbol{\beta}_k + \mathbf{Y}_i^T \boldsymbol{\alpha}_k + \boldsymbol{\eta}^T \boldsymbol{\phi}_k} \Lambda_k(\tilde{T}_{ki}) \right\} \right] \\ &\quad \times f_Y(\mathbf{Y}_i \mid \mathbf{Z}_i, \boldsymbol{\eta}; \boldsymbol{\psi}) f_{\boldsymbol{\eta}}(\boldsymbol{\eta} \mid \mathbf{Z}_i; \boldsymbol{\nu}) d\boldsymbol{\eta}, \end{aligned}$$

$\dot{\Psi}_{\boldsymbol{\theta}}(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A})$  be the derivative of  $\Psi(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A})$  with respect to  $\boldsymbol{\theta}$ , and  $\dot{\Psi}_k(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A})[H_k]$  be the derivative of  $\Psi(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A})$  along the path  $(\Lambda_k + \epsilon H_k)$ .

First, we prove the consistency. By condition (D4),

$$\begin{aligned} &\left[ e^{\mathbf{Z}_i^T \boldsymbol{\beta}_k + \mathbf{Y}_i^T \boldsymbol{\alpha}_k + \boldsymbol{\eta}^T \boldsymbol{\phi}_k} G'_k \left\{ e^{\mathbf{Z}_i^T \boldsymbol{\beta}_k + \mathbf{Y}_i^T \boldsymbol{\alpha}_k + \boldsymbol{\eta}^T \boldsymbol{\phi}_k} \Lambda_k(\tilde{T}_{ki}) \right\} \right]^{\Delta_{ki}} \\ &\quad \times e^{-G_k \left\{ e^{\mathbf{Z}_i^T \boldsymbol{\beta}_k + \mathbf{Y}_i^T \boldsymbol{\alpha}_k + \boldsymbol{\eta}^T \boldsymbol{\phi}_k} \Lambda_k(\tilde{T}_{ki}) \right\}} \\ &\leq e^{O(1+|\mathbf{Y}|+|\boldsymbol{\eta}|)} \{1 + \Lambda_k(\tilde{T}_{ki})\}^{-\Delta_{ki} - \kappa_{3k} + \kappa_{2k} + 1}. \end{aligned}$$

Thus, condition (D3) implies that

$$\Psi(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A}) \leq \prod_{i=1}^n \mathcal{F}(\mathcal{O}_i; \boldsymbol{\theta}) \prod_{k=1}^K \left\{ 1 + \Lambda_k(\tilde{T}_{ki}) \right\}^{-\Delta_{ki} - \kappa_{3k} + \kappa_{2k} + 1}, \quad (\text{A.5})$$

where  $\mathcal{F}(\mathcal{O}_i; \boldsymbol{\theta})$  is a random variable with  $|\mathbb{E}[\log \mathcal{F}(\mathcal{O}_i; \boldsymbol{\theta})]| < \infty$  for any  $\boldsymbol{\theta}$ . By condition (D2),  $P(\tilde{T}_{ki} = \tau)$  is positive. Therefore, if  $\Lambda_k(\tau) = \infty$ , then the right-hand side of (A.5) is zero for large  $n$ . We conclude that  $\hat{\Lambda}_k(\tau) < \infty$ , such that the NPMLE exists.

We then show that  $\limsup_n \hat{\Lambda}_k(\tau) < \infty$  almost surely. From (A.5),

$$\begin{aligned} \frac{1}{n} \log L_n(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) &= \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \Delta_{ki} \log \hat{\Lambda}_k(\tilde{T}_{ki}) + \frac{1}{n} \sum_{i=1}^n \log \Psi(\mathcal{O}_i; \hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) \\ &\leq \frac{1}{n} \sum_{i=1}^n \log \mathcal{F}(\mathcal{O}_i; \hat{\boldsymbol{\theta}}) + \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \Delta_{ki} \log \hat{\Lambda}_k(\tilde{T}_{ki}) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K (\Delta_{ki} + \kappa_{3k} - \kappa_{2k} - 1) \log \left\{ 1 + \hat{\Lambda}_k(\tilde{T}_{ki}) \right\}. \end{aligned}$$

Let  $\tilde{N} = n^{-1} \sum_{i=1}^n (\Delta_{1i} I(\tilde{T}_{1i} \leq \cdot), \dots, \Delta_{Ki} I(\tilde{T}_{Ki} \leq \cdot))$ . Clearly,

$$\frac{1}{n} \log L_n(\boldsymbol{\theta}_0, \tilde{N}) = -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \Delta_{ki} \log n + \frac{1}{n} \sum_{i=1}^n \log \Psi(\mathcal{O}_i; \boldsymbol{\theta}_0, \tilde{N}).$$

The second term on the right-hand side of the above equation is  $O_p(1)$ . Thus,

$$\begin{aligned} &\frac{1}{n} \log L_n(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) - \frac{1}{n} \log L_n(\boldsymbol{\theta}_0, \tilde{N}) + O_p(1) \\ &\leq \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \Delta_{ki} \log [n \hat{\Lambda}_k(\tilde{T}_{ki})] \\ &\quad - \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K (\Delta_{ki} + \kappa_{3k} - \kappa_{2k} - 1) \log \left\{ 1 + \hat{\Lambda}_k(\tilde{T}_{ki}) \right\}. \end{aligned}$$

Note that  $(\kappa_{3k} - \kappa_{2k} - 1)$  is positive by condition (D4). Using the partitioning argument similar to those of Murphy (1994) and

Parner (1998), we can show that the right-hand side of the above inequality tends to  $-\infty$  if  $\limsup_n \hat{\Lambda}_k(\tau) = \infty$ . By definition of  $(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})$ , the left-hand side of the inequality is bounded below by an  $O_p(1)$  term. Therefore,  $\hat{\Lambda}_k(\tau)$  is uniformly bounded.

Given the boundedness of  $\hat{\Lambda}_k(\tau)$ , Helly's selection theorem implies that, for any subsequence of  $n$ , we can always choose a further subsequence such that  $\hat{\Lambda}_k$  converges pointwise to some monotone function  $\Lambda_k^*$  and  $\hat{\boldsymbol{\theta}}$  converges to  $\boldsymbol{\theta}^*$ . The desired consistency result follows if we can show that  $\Lambda_k^* = \Lambda_{0k}$  and  $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$  almost surely. With an abuse of notation, let  $\{n\}_{1,2,\dots}$  be the subsequence. Define

$$\tilde{\Lambda}_k(t) = - \sum_{i=1}^n \Delta_{ki} I(\tilde{T}_{ki} \leq t) \left\{ \sum_{j=1}^n \frac{\dot{\Psi}_k(\mathcal{O}_j; \boldsymbol{\theta}_0, \mathcal{A}_0)[I(\tilde{T}_{ki} \leq \cdot)]}{\Psi(\mathcal{O}_j; \boldsymbol{\theta}_0, \mathcal{A}_0)} \right\}^{-1}.$$

By Lemma S2 in the Supplementary Materials and the properties of Donsker (and therefore, Glivenko-Cantelli) classes,

$$\frac{1}{n} \sum_{j=1}^n \frac{\dot{\Psi}_k(\mathcal{O}_j; \boldsymbol{\theta}_0, \mathcal{A}_0)[I(s \leq \cdot)]}{\Psi(\mathcal{O}_j; \boldsymbol{\theta}_0, \mathcal{A}_0)} \rightarrow \mathbb{E} \left( \frac{\dot{\Psi}_k(\mathcal{O}_i; \boldsymbol{\theta}_0, \mathcal{A}_0)[I(s \leq \cdot)]}{\Psi(\mathcal{O}_i; \boldsymbol{\theta}_0, \mathcal{A}_0)} \right)$$

uniformly on  $[0, \tau]$ . Because the score function along the path  $\Lambda_k = \Lambda_{0k} + \epsilon I(\cdot \geq s)$  with other parameters fixed at their true values has zero expectation,

$$-\mathbb{E} \left\{ \frac{\dot{\Psi}_k(\mathcal{O}_i; \boldsymbol{\theta}_0, \mathcal{A}_0)[I(s \leq \cdot)]}{\Psi(\mathcal{O}_i; \boldsymbol{\theta}_0, \mathcal{A}_0)} \right\} = \frac{dP(\tilde{T}_{ki} \Delta_{ki} < s)/ds}{\lambda_{0k}(s)}.$$

Algebraic manipulation yields that the uniform limit of  $\tilde{\Lambda}_k$  on  $[0, \tau]$  is  $\Lambda_{0k}$ . Note that  $\hat{\Lambda}_k(t)$  is equal to

$$\int_0^t \frac{\left| n^{-1} \sum_{j=1}^n \dot{\Psi}_k(\mathcal{O}_j; \boldsymbol{\theta}_0, \mathcal{A}_0)[I(s \leq \cdot)] / \Psi(\mathcal{O}_j; \boldsymbol{\theta}_0, \mathcal{A}_0) \right|}{\left| n^{-1} \sum_{j=1}^n \dot{\Psi}_k(\mathcal{O}_j; \hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})[I(s \leq \cdot)] / \Psi(\mathcal{O}_j; \hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) \right|} d\tilde{\Lambda}_k(s).$$

We have shown that the numerator of the integrand in the above equation converges uniformly. Similarly, we can show that the denominator of the integrand in the above equation converges uniformly to  $|\mathbb{E}\{\dot{\Psi}_k(\mathcal{O}_i; \boldsymbol{\theta}^*, \mathcal{A}^*)[I(s \leq \cdot)] / \Psi(\mathcal{O}_i; \boldsymbol{\theta}^*, \mathcal{A}^*)\}|$  and that the limit is bounded away from 0. Because  $\tilde{\Lambda}_k$  converges uniformly to  $\Lambda_{0k}$ , which is differentiable with respect to  $t$ ,  $\Lambda_k^*$  is also differentiable with respect to  $t$ . It follows that  $d\hat{\Lambda}_k/d\tilde{\Lambda}_k$  converges uniformly to  $\lambda_k^*/\lambda_{0k}$  on  $[0, \tau]$ , where  $\lambda_k^* = (\Lambda_k^*)'$ . As  $n^{-1} \log L_n(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) - n^{-1} \log L_n(\boldsymbol{\theta}_0, \tilde{N})$  is nonnegative,

$$\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \Delta_{ki} \log \frac{d\hat{\Lambda}_k(\tilde{T}_{ki})}{d\tilde{\Lambda}_k(\tilde{T}_{ki})} + \frac{1}{n} \sum_{i=1}^n \log \frac{\Psi(\mathcal{O}_i; \hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})}{\Psi(\mathcal{O}_i; \boldsymbol{\theta}_0, \tilde{N})} \geq 0.$$

By the Glivenko-Cantelli properties of the class of functions of  $\log \Psi(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A})$  given by Lemma S2 and the uniform convergence of  $d\hat{\Lambda}_k/d\tilde{\Lambda}_k$ , setting  $n \rightarrow \infty$  on both sides of the above inequality yields

$$\mathbb{E} \left\{ \log \frac{\prod_{k=1}^K \lambda_k^*(\tilde{T}_{ki})^{\Delta_{ki}} \Psi(\mathcal{O}_i; \boldsymbol{\theta}^*, \mathcal{A}^*)}{\prod_{k=1}^K \lambda_{0k}(\tilde{T}_{ki})^{\Delta_{ki}} \Psi(\mathcal{O}_i; \boldsymbol{\theta}_0, \mathcal{A}_0)} \right\} \geq 0.$$

The left-hand side of the above inequality is the negative Kullback-Leibler distance of the density indexed by  $(\theta^*, \mathcal{A}^*)$ . From the identifiability of the model implied by Theorem 1, we conclude that  $\theta^* = \theta_0$  and  $\Lambda_k^* = \Lambda_{k0}$ . The desired consistency result follows.

To prove the asymptotic normality of the NPMLE, we adopt the arguments of van der Vaart (1998, pp. 419–424). Let  $\mathcal{P}_n$  be the empirical measure determined by  $n$  iid observations, and let  $\mathcal{P}$  be the true probability measure. Let  $\dot{\ell}_\theta(\theta, \mathcal{A})$  be the derivative of  $\log L_n(\theta, \mathcal{A})$  with respect to  $\theta$ , and let  $\dot{\ell}_k(\theta, \mathcal{A})[H_k]$  be the derivative of  $\log L_n(\theta, \mathcal{A})$  along the path  $(\Lambda_k + \epsilon H_k)$ . For any  $\mathbf{v} \in \mathbb{R}^d$  and  $\mathcal{W} = (h_1, \dots, h_K)$  with  $h_k \in \text{BV}[0, \tau]$ , where  $\text{BV}[0, \tau]$  is the space of functions of bounded variation on  $[0, \tau]$ , we have

$$\mathcal{P}_n \left( \mathbf{v}^T \dot{\ell}_\theta(\hat{\theta}, \hat{\mathcal{A}}) + \sum_{k=1}^K \dot{\ell}_k(\hat{\theta}, \hat{\mathcal{A}}) \left[ \int h_k d\hat{\Lambda}_k \right] \right) = 0.$$

In addition,

$$\mathcal{P} \left( \mathbf{v}^T \dot{\ell}_\theta(\theta_0, \mathcal{A}_0) + \sum_{k=1}^K \dot{\ell}_k(\theta_0, \mathcal{A}_0) \left[ \int h_k d\Lambda_{0k} \right] \right) = 0.$$

Therefore,

$$\begin{aligned} & \sqrt{n}(\mathcal{P}_n - \mathcal{P}) \left( \mathbf{v}^T \dot{\ell}_\theta(\hat{\theta}, \hat{\mathcal{A}}) + \sum_{k=1}^K \dot{\ell}_k(\hat{\theta}, \hat{\mathcal{A}}) \left[ \int h_k d\hat{\Lambda}_k \right] \right) \\ &= -\sqrt{n} \mathcal{P} \left\{ \left( \mathbf{v}^T \dot{\ell}_\theta(\hat{\theta}, \hat{\mathcal{A}}) + \sum_{k=1}^K \dot{\ell}_k(\hat{\theta}, \hat{\mathcal{A}}) \left[ \int h_k d\hat{\Lambda}_k \right] \right) \right. \\ & \quad \left. - \left( \mathbf{v}^T \dot{\ell}_\theta(\theta_0, \mathcal{A}_0) + \sum_{k=1}^K \dot{\ell}_k(\theta_0, \mathcal{A}_0) \left[ \int h_k d\Lambda_{0k} \right] \right) \right\}. \end{aligned} \quad (\text{A.6})$$

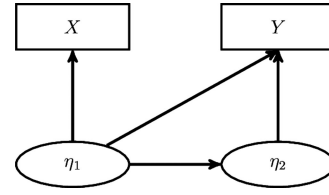
From the Donsker properties of the classes of functions of  $\dot{\ell}_\theta$  and  $\dot{\ell}_k$  implied by Lemma S2 and the consistency of  $\hat{\theta}$  and  $\hat{\mathcal{A}}$ , we conclude that the left-hand side of (A.6) equals

$$\begin{aligned} & \sqrt{n}(\mathcal{P}_n - \mathcal{P}) \left( \mathbf{v}^T \dot{\ell}_\theta(\theta_0, \mathcal{A}_0) + \sum_{k=1}^K \dot{\ell}_k(\theta_0, \mathcal{A}_0) \left[ \int h_k d\Lambda_{0k} \right] \right) \\ & + o_p(1). \end{aligned}$$

This term converges to a Gaussian process in  $l^\infty(\mathcal{V} \times \mathcal{Q}^K)$ . By the Taylor series expansion, the right-hand side of (A.6) is of the form

$$\begin{aligned} & -\sqrt{n} \left\{ B_1[\mathbf{v}, \mathcal{W}]^T (\hat{\theta} - \theta_0) + \sum_{k=1}^K \int B_{2k}[\mathbf{v}, \mathcal{W}] d(\hat{\Lambda}_k - \Lambda_{0k}) \right\} \\ & + o_p \left( \sqrt{n} \|\hat{\theta} - \theta_0\| + \sqrt{n} \sum_{k=1}^K \left\| \hat{\Lambda}_k - \Lambda_{0k} \right\|_{V[0, \tau]} \right), \end{aligned}$$

where  $\mathcal{B} \equiv (B_1, B_{21}, \dots, B_{2K})$  is the information operator and is linear in  $\mathbb{R}^d \times \text{BV}[0, \tau]^K$ . By Lemma 2,  $\mathcal{B}$  is invertible. The rest of the proof then follows the arguments of van der Vaart (1998, pp. 419–424). Finally, because  $\mathbf{v}^T \hat{\theta}$  is an asymptotically linear estimator of  $\mathbf{v}^T \theta_0$  with the influence function lying in the space spanned by the score functions,  $\hat{\theta}$  is an efficient estimator for  $\theta_0$ .  $\square$



**Figure A.1.** SEM considered in Lemma 1. The SEM consists of two sets of latent variables and two sets of observed variables that may all be multivariate. The observed variable  $X$  depends only on the latent variable  $\eta_1$ , but the observed variable  $Y$  depends on both sets of latent variables.

The following two lemmas are used in the proofs of Theorem 1 and Theorem 2 and are proved in Section S2 of the Supplementary Materials.

*Lemma A.1.* Let Model A be

$$\begin{aligned} X | (\eta_1, \eta_2) &\stackrel{d}{=} X | \eta_1 \sim F_{X|\eta_1}(\cdot | \eta_1), \\ Y | (\eta_1, \eta_2) &\sim F_{Y|\eta}(\cdot | \eta_1, \eta_2), \\ \eta_2 | \eta_1 &\sim F_{\eta_2|\eta_1}(\cdot | \eta_1), \\ \eta_1 &\sim F_{\eta_1}, \end{aligned}$$

where  $(X, Y)$  are observed, and  $(\eta_1, \eta_2)$  are latent. Model A is depicted in Figure A.1. Let  $f_{X|\eta_1} = F'_{X|\eta_1}$ ,  $f_{Y|\eta} = F'_{Y|\eta}$ ,  $f_{\eta_2|\eta_1} = F'_{\eta_2|\eta_1}$ , and  $f_{\eta_1} = F'_{\eta_1}$ . Assume that: (a) for any density functions  $\tilde{f}_{Y|\eta}$  and  $\tilde{f}_{\eta_2|\eta_1}$ ,

$$\begin{aligned} & \int f_{Y|\eta}(Y | \eta_1, \eta_2) f_{\eta_2|\eta_1}(\eta_2 | \eta_1) d\eta_2 \\ &= \int \tilde{f}_{Y|\eta}(Y | \eta_1, \eta_2) \tilde{f}_{\eta_2|\eta_1}(\eta_2 | \eta_1) d\eta_2 \quad \forall Y, \eta_1 \end{aligned}$$

implies that  $(f_{Y|\eta}, f_{\eta_2|\eta_1}) = (\tilde{f}_{Y|\eta}, \tilde{f}_{\eta_2|\eta_1})$ , that is, the model for  $Y$  is identifiable if  $\eta_1$  is observed; (b)  $F_{X|\eta_1}$  and  $F_{\eta_1}$  are identifiable based on  $(X, Y)$ ; and (c)  $\eta_1$  is a complete sufficient statistic in  $\{F_{\eta_1|X}(\cdot | X) : X \in \mathcal{X}\}$ , where  $F_{\eta_1|X}$  is the conditional distribution function of  $\eta_1$  given  $X$ , and  $\mathcal{X}$  is the range of  $X$ . Then, Model A is identifiable. A sufficient condition for  $\eta_1$  to be complete sufficient is that the density of  $X$  is of the form

$$f_{X|\eta_1}(X | \eta_1) \propto \prod_{j=1}^q \exp \{X_j s_j(\eta_1) - a_j(\eta_1)\} b_j(X_j),$$

where  $\mathbf{X} = (X_1, \dots, X_q)$ ,  $\eta_1 \mapsto (s_1(\eta_1), \dots, s_q(\eta_1))$  is one-to-one, and  $b_j$  is nonzero on some open set.

*Lemma 2.* Under conditions (C1), (C2), (C3'), (C5'), and (D5), the model given by (1)–(3) has an invertible information operator.

## Supplementary Materials

The supplementary materials contain proofs of Lemmas 1 and 2, additional theoretical results, simulation results for Mplus, and additional data analysis results.

## Acknowledgments

Kin Yau Wong is Doctoral Student (E-mail: alexwky@live.unc.edu), Donglin Zeng is Professor (E-mail: dzeng@bios.unc.edu), and D. Y. Lin is Dennis Gillings Distinguished Professor (E-mail: lin@bios.unc.edu).



edu), Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA.

## Funding

This research was supported by the National Institutes of Health grants R01GM047845, R01CA082659, and P01CA142538.

## References

- Asparouhov, T., Masyn, K., and Muthén, B. (2006), “Continuous Time Survival in Latent Variable Models,” in *Proceedings of the Joint Statistical Meeting in Seattle, August 2006. ASA Section on Biometrics*, pp. 180–187. [893]
- Bollen, K. A. (1989), *Structural Equations With Latent Variables*, New York: Wiley. [893,894,895]
- Bollen, K. A., and Davis, W. R. (2009), “Two Rules of Identification for Structural Equation Models,” *Structural Equation Modeling*, 16, 523–536. [894]
- Chareka, P. (2007), “A Finite-Interval Uniqueness Theorem for Bilateral Laplace Transforms,” *International Journal of Mathematics and Mathematical Sciences*, 2007. [902]
- Cox, D. R. (1972), “Regression Models and Life-Tables,” *Journal of the Royal Statistical Society, Series B*, 34, 187–220. [893]
- Dahly, D., Adair, L., and Bollen, K. (2009), “A Structural Equation Model of the Developmental Origins of Blood Pressure,” *International Journal of Epidemiology*, 38, 538–548. [893]
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), “Maximum Likelihood From Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society, Series B*, 39, 1–38. [895]
- Henderson, R., Diggle, P., and Dobson, A. (2000), “Joint Modelling of Longitudinal Measurements and Event Time Data,” *Biostatistics*, 1, 465–480. [893]
- Huang, J., and Wellner, J. A. (1997), “Interval Censored Survival Data: A Review of Recent Progress,” in *Proceedings of the First Seattle Symposium in Biostatistics*, eds. D. Y. Lin, and T. R. Fleming, New York: Springer, pp. 123–169. [901]
- Kortram, R. A., van Rooij, A. C. M., Lenstra, A. J., and Ridder, G. (1995), “Constructive Identification of the Mixed Proportional Hazards Model,” *Statistica Neerlandica*, 49, 269–281. [901]
- Kosorok, M. R., Lee, B. L., and Fine, J. P. (2004), “Robust Inference for Univariate Proportional Hazards Frailty Regression Models,” *The Annals of Statistics*, 32, 1448–1491. [894,898]
- Larsen, K. (2004), “Joint Analysis of Time-to-Event and Multiple Binary Indicators of Latent Classes,” *Biometrics*, 60, 85–92. [893]
- (2005), “The Cox Proportional Hazards Model With a Continuous Latent Variable Measured by Multiple Binary Indicators,” *Biometrics*, 61, 1049–1055. [893]
- Liu, Q., and Pierce, D. A. (1994), “A Note on Gauss-Hermite Quadrature,” *Biometrika*, 81, 624–629. [896]
- Louis, T. A. (1982), “Finding the Observed Information Matrix When Using the EM Algorithm,” *Journal of the Royal Statistical Society, Series B*, 44, 226–233. [896]
- Menendez, J. A., and Lupu, R. (2007), “Fatty Acid Synthase and the Lipogenic Phenotype in Cancer Pathogenesis,” *Nature Reviews Cancer*, 7, 763–777. [900]
- Moustaki, I., and Steele, F. (2005), “Latent Variable Models for Mixed Categorical and Survival Responses, With an Application to Fertility Preferences and Family Planning in Bangladesh,” *Statistical Modelling*, 5, 327–342. [893]
- Murphy, S. A. (1994), “Consistency in a Proportional Hazards Model Incorporating a Random Effect,” *The Annals of Statistics*, 22, 712–731. [903]
- Muthén, B., and Masyn, K. (2005), “Discrete-Time Survival Mixture Analysis,” *Journal of Educational and Behavioral Statistics*, 30, 27–58. [893]
- Muthén, L., and Muthén, B. (1998–2015), *Mplus User’s Guide, 7th ed.*, Los Angeles, CA: Muthén & Muthén. [893,898]
- Naliboff, B. D., Kim, S. E., Bolus, R., Bernstein, C. N., Mayer, E. A., and Chang, L. (2012), “Gastrointestinal and Psychological Mediators of Health-Related Quality of Life in IBS and IBD: A Structural Equation Modeling Analysis,” *The American Journal of Gastroenterology*, 107, 451–459. [893]
- Parner, E. (1998), “Asymptotic Theory for the Correlated Gamma-Frailty Model,” *The Annals of Statistics*, 26, 183–214. [903]
- Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004), “Generalized Multilevel Structural Equation Modeling,” *Psychometrika*, 69, 167–190. [893]
- Rabe-Hesketh, S., Yang, S., and Pickles, A. (2001), “Multilevel Models for Censored and Latent Responses,” *Statistical Methods in Medical Research*, 10, 409–427. [893]
- Reilly, T., and O’Brien, R. M. (1996), “Identification of Confirmatory Factor Analysis Models of Arbitrary Complexity: The Side-By-Side Rule,” *Sociological Methods & Research*, 24, 473–491. [894,897]
- Stoolmiller, M., and Snyder, J. (2006), “Modeling Heterogeneity in Social Interaction Processes Using Multilevel Survival Analysis,” *Psychological Methods*, 11, 164–177. [893]
- Stoolmiller, M., and Snyder, J. (2014), “Embedding Multilevel Survival Analysis of Dyadic Social Interaction in Structural Equation Models: Hazard Rates as Both Outcomes and Predictors,” *Journal of Pediatric Psychology*, 39, 222–232. [893]
- The Cancer Genome Atlas Research Network, (2011), “Integrated Genomic Analyses of Ovarian Carcinoma,” *Nature*, 474, 609–615. [899]
- Tibshirani, R. (1996), “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society, Series B*, 58, 267–288. [901]
- Tsiatis, A. A., and Davidian, M. (2004), “Joint Modeling of Longitudinal and Time-To-Event Data: An Overview,” *Statistica Sinica*, 14, 809–834. [893]
- van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge: Cambridge University Press. [902,904]
- Vicard, P. (2000), “On the Identification of a Single-Factor Model With Correlated Residuals,” *Biometrika*, 87, 199–205. [894]
- Yuan, M., and Lin, Y. (2006), “Model Selection and Estimation in Regression With Grouped Variables,” *Journal of the Royal Statistical Society, Series B*, 68, 49–67. [901]
- Zeng, D., and Lin, D. Y. (2010), “A General Asymptotic Theory for Maximum Likelihood Estimation in Semiparametric Regression Models With Censored Data,” *Statistica Sinica*, 20, 871–910. [894,897,898,901]