1. INTRODUCTION

The proportional and the additive hazards (AH) models are two popular frameworks in biomedical studies for investigating the association between risk factors and disease occurrence or death (Cox 1972; Cox and Oakes 1984; Aalen 1989; Huffer and McKeague 1991; Lin and Ying 1994). The AH model provides a characterization of the effects of risk factors different from the proportional hazards (PH) model, and has remarkable features that are not found in the latter model. In particular, the AH model pertains to the risk difference, which is especially relevant and informative in epidemiological and clinical studies. Similar to the PH model, the AH model also has sound biological and empirical bases (Breslow and Day 1987).

There are many situations in medical, behavioral, and psychological research settings when outcomes or potential risk factors cannot be measured by a single observed variable. Instead, they should be characterized by several observed variables from different perspectives. Examples include glycemia measured by glycated hemoglobin (HbA1c) and fasting plasma glucose (FPG); lipid characterized by total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TG); and behavioral problem assessed by antisocial, anxious, dependent, headstrong, and hyperactive behavior. The conventional PH and AH models manage latent factors by incorporating one or more of their surrogates into the regression analysis. However, problems may exist with the separate inclusion of partial surrogates of the latent factors. First, important attributes of latent factors may be lost because of incomplete measurement, leading to bias or misleading results. For instance, systolic blood pressure (SBP) and diastolic blood pressure (DBP) are usually used together to characterize blood pressure. Although highly correlated, SBP and DBP reflect blood pressure from different angles and do not replace each other. If one only uses SBP (or DBP) to measure blood pressure, then the insufficient information cannot reveal the danger of too high (low) DBP (or SBP), leading to incorrect conclusion for people with normal SBP (or DBP) but abnormal DBP (or SBP). Second, multiple surrogates of a latent factor are highly correlated because they share certain similarity elicited by this latent factor. Such high correlation may cause the multicollinearity problem, resulting in incorrect estimation or divergence of the computer program. The latter example (see Table 4) shows that incorporating the individual highly correlated observed variables WAIST, BMI, SBP, DBP, HbA1c, FPG, TC, HDL-C, and TG in a standard analysis causes multicollinearity and produces confounding results. Finally, interpreting the effect of each surrogate separately may be tedious and incomprehensible. For instance, when examining the effect of lipid control on the development of chronic kidney disease (CKD), a standard analysis only tells how the individual observed variables TC, HDL-C, and TG influence the hazards of CKD but never addresses the major concern on the overall impact of lipid on CKD. Further, the diverse effects caused by the multicollinearity (Table 4) makes the interpretation even harder. By contrast, the proposed joint analysis groups TC, HDL-C, and TG together into a latent variable "lipid" and provides scientific evidence for the overall effect of this latent variable and thus makes the interpretation conceptually simple and comprehensible. Simultaneously, the dimension of covariates is also reduced substantially.

A number of studies have assessed the interdependence, causal relationships, and correlations among multiple observed variables and latent factors. The commonly used techniques include factor analysis (Lawley and Maxwell 1971) and latent variable models (Bollen 1989; Jöreskog and Sörbom 1996; Lee 2007; Song and Lee 2012). The latent variable analysis originated in the field of psychology to relate various mental test results to latent
traits, such as intelligence. Additionally, it has recently been used in medical studies to investigate the relationships between covariates (e.g., age) and a medical trait (e.g., overall treatment effect), or between the endpoint of interest (e.g., kidney failure) and various potential latent risk factors (e.g., glycemia and lipid). A typical modeling framework for such latent variable analysis consists of two models. The first model is the so-called confirmatory factor analysis (CFA) model (Bollen 1989, chap. 7; Shi and Lee 2000; Lee 2007, chap. 2) for characterizing latent factors through multiple correlated observed variables, whereas the other one is a regression type model for relating observed and latent risk factors to quantities of interest. Sammel and Ryan (1996) measured the overall severity of birth defects using multiple adverse effects and then regressed it on the exposure to anticonvulsants when analyzing the effects of anticonvulsant medication during pregnancy. Roy and Lin (2000, 2002) summarized the effectiveness of treatment practices using three longitudinal outcomes and related the overall treatment effect to covariates through a linear mixed model when examining the effectiveness of methadone treatment practice in reducing illicit drug use.

In this study, we propose a joint model to analyze the survival data with latent variables. The proposed model consists of two components: a CFA model for characterizing the latent risk factors, and an AH model to relate the observed and latent risk factors to the hazard function of interest. To analyze such kinds of joint models with latent variables, published reports (e.g., Sammel and Ryan 1996; Roy and Lin 2002) have mainly used the expectation–maximization (EM) algorithm, in which the latent variables and/or random effects were treated as missing data, and the maximum likelihood estimates (MLEs) of the model parameters were obtained by maximizing the expectation of the complete-data log-likelihood function. In our proposed model, however, the AH model is nonlinear with an unspecified baseline hazard function, which makes the complete-data log-likelihood function highly intractable. Directly applying the EM algorithm to analyze the proposed model is challenging. Thus, to address such difficulty, we develop a borrow-strength estimation procedure in this study. The procedure is a two-stage approach. At the first stage, latent variables are characterized by the CFA model and estimated using the EM algorithm. At the second stage, the AH model with the estimated latent variables is analyzed using the corrected pseudo-score method (Carroll, Ruppert, and Stefanski 1995; Kulich and Lin 2000; Song and Huang 2006). To the best of our knowledge, no study has analyzed the proposed joint survival model with latent variables.

The present research is motivated by a study based on the Hong Kong Diabetes Registry, which was established in 1995 as part of a continuous quality improvement program at the Prince of Wales Hospital in Hong Kong. A 10-year prospective cohort of 3586 Chinese Type 2 diabetic patients who may have experienced CKD, one of the most common diabetic complications, was analyzed. A primary interest of this study is to investigate the potential risk factors of CKD to prevent diabetic complications and improve the quality of life for diabetic patients. However, some risk factors, such as obesity, blood pressure, glycemia, and lipid, are latent traits that cannot be measured by a single observed variable. Thus, how these latent risk factors affect the overall development of CKD is of great interest in medical research. To address this problem, we propose the current joint model with latent variables. The empirical analysis (Section 5) shows that our model provides important insights into the overall effects of latent risk factors. Such kinds of overall effects cannot be revealed by conventional AH models.

The rest of the article is organized as follows. In Section 2, we describe the proposed joint model and present the borrow-strength estimation procedure for regression parameters of interest. The asymptotic properties of the proposed estimators are established in Section 3. Section 4 presents simulation studies to evaluate the empirical performance of the proposed methods. An application to the CKD data is reported in Section 5. Section 6 concludes the article with discussion. Proofs are provided in the Appendix and other details are given in supplementary material.

2. MODEL AND INFERENC PROCEDURES

2.1 Joint Models

We propose a joint model that involves two major components. The first component is the CFA model and the second component is the AH model. For \( i = 1, \ldots, n \), let \( \mathbf{V}_i \) be a \( p \times 1 \) vector of the observed variables and \( \xi_i \) be a \( q \times 1 \) vector of latent variables (factors), where \( p > q \). The CFA model for characterizing latent variables through multiple observed variables is defined as

\[
\mathbf{V}_i = \mathbf{B} \xi_i + \mathbf{e}_i,
\]

where \( \mathbf{B} \) is a \( p \times q \) matrix of factor loadings, \( \xi_i \sim N(0, \Phi) \), and \( \mathbf{e}_i \) is a \( p \times 1 \) vector of random errors independent of \( \xi_i \) and distributed as \( N(0, \Psi_e) \) with a diagonal \( \Psi_e \). The elements of \( \mathbf{B} \) reflect the associations between the observed variables and their corresponding latent variables. In a CFA model, any element of \( \mathbf{B} \) can be fixed at preassigned values. This setup allows great flexibility in selecting an appropriate structure of \( \mathbf{B} \) to achieve a clear interpretation of the latent variables. In substantive research, we usually already have a clear idea of the number of latent variables and the structure of the factor loading matrix. For example, in the CKD study, obesity, blood pressure, glycemia, and lipid are important medical traits for assessing the health profile of diabetic patients. Our main objective is to examine their effects on the risk of CKD. Thus, the number of latent variables in this study is clearly four. Experts in diabetes and the existing medical literature (e.g., Song et al. 2009; Wang et al. 2014) suggested appropriate observed variables for these latent variables. Thus, the number of latent variables and the structure of \( \mathbf{B} \) can be naturally decided based on the study objectives, the meaning of the observed variables suggested by subject experts, and/or the existing literature. Moreover, an exploratory factor analysis (EFA) can be used to cross validate the decision. Well-known software in structural equation modeling, such as LISREL (Jöreskog and Sörbom 1996), EQS (Bentler 2002), and Mplus (Muthén and Muthén 1998–2007), can be used to conduct the EFA. In the analysis of CKD data, we conducted an EFA using LISREL to verify the decision on the number of factors and the structure of \( \mathbf{B} \) (see details in Section 5).

To investigate the potential risk factors of a failure time of interest, the latent variables characterized by the CFA model, along with some observed covariates, are further related to the
failure time through an AH model with latent variables. Let \( T_i \) be the failure time of interest, \( Z_i \) be a \( s \times 1 \) vector of covariates, and \( C_i \) be the censoring time assumed to be conditionally independent of \( T_i \) given \( Z_i \) and \( \xi_i \). Let \( X_i = \min(T_i, C_i) \) be the observed time and \( \Delta_i = I(T_i \leq C_i) \) be the failure indicator, where \( I(\cdot) \) is the indicator function. The AH model specifies that, given \( Z_i \) and \( \xi_i \), the hazard function of \( T_i \) takes the form
\[
\lambda(t | Z_i, \xi_i) = \lambda_0(t) + \beta^T Z_i + \gamma^T \xi_i, \tag{2}
\]
where \( \lambda_0(t) \) is an unspecified baseline hazard function, and \( \beta \) and \( \gamma \) are the \( s \times 1 \) and \( q \times 1 \) vectors of unknown regression parameters. Given that \( \xi_i \) plays a role of risk factors in model (2), the observed variables in model (1) are actually multiple risk factors that are related to a latent variable (another risk factor). Thus, the proposed joint model is different from a typical joint model literature, in which several post-exposure outcome variables of interest exist.

Song and Huang (2006) developed a joint analysis to accommodate longitudinal covariates measured with error in an AH model. Their first model is essentially a growth curve model with measurement error. It evaluates a time-dependent covariate through a single surrogate with measurement error at intermittent occasions. By contrast, our first model emphasizes the use of “latent variables” to assess latent traits that should be measured by several (usually highly correlated) observed variables. For example, the latent variable “blood pressure” is measured by observed variables SBP and DBP, and the latent variable “lipid” is measured by observed variables TC, HDL-C, and TG. This kind of latent variables contributes an integral part in structural equation models and has been widely used in the medical, behavioral, and social sciences (see Bollen 1989; Lee 2007; Song and Lee 2012, and references therein). Our first model is a CFA model that is used to form the latent variables in \( \xi_i \) via grouping appropriate observed variables through a factor loading matrix \( B \). In the CFA analysis, the unknown parameters in \( B \), which can be regarded as data-driven weights to reflect how the observed variables contribute to the characterization of a latent variable, are estimated along with the correlations of the latent variables and the error variances. Clearly, Song and Huang’s model and other measurement error models do not involve such kind of latent variables and the factor loading matrix. Thus, the CFA model in (1) is quite different from the growth curve measurement error model of Song and Huang (2006). In formulating the AH model, Song and Huang (2006) directly incorporated measurement error covariates into the model. By contrast, we propose the use of latent variables obtained from the CFA, rather than a large number of covariates, to model the hazard function of interest.

While our model and that of Song and Huang (2006) are different in terms of scientific goal, modeling of latent variables, and scope of applications, they share certain similarities. First, the method proposed by Song and Huang (2006) assumed mixed effects model for time-dependent covariates. Time-independent covariates with repeated measurements can be viewed as a special case with the slope equal to 0. Thus, when the \( B \) matrix is block diagonal with each block being all 1’s vector, the CFA model can be viewed similarly as a measurement error model. Second, given that both methods involve variables measured with errors, they encounter similar difficulty induced by unobservable variables and bear the same spirit of using corrected pseudo-score approach in statistical inference.

The proposed model subsumes several statistical models as its special cases. For instance, if \( p = q = 0 \), then the proposed model reduces to a simple AH model with only fixed effects. If \( p = q \) and \( B = I \) (identity matrix), then the proposed model reduces to an AH model with covariates measured with error. If one only considers model (2) and fixes the elements of \( \gamma \) at 1, then the proposed model reduces to a frailty AH model with both fixed and random effects. Compared with the conventional AH model or frailty AH models (Yin and Ibrahim 2005; Cai and Zeng 2011), the proposed joint model has several appealing features. First, the analytic power of the joint model is greatly enhanced using the information from multiple indicators. Second, the joint model provides statistical evidence and attractive interpretation for the effects of latent traits that are well known but hard to measure directly. The frailty terms in conventional AH models only address the dependence among clustered or longitudinal observations, whereas the latent variables in the proposed model represent medical traits with specific meanings. These latent variables not only address the possible heterogeneity in the data but also provide insights into their impacts on the failure time of interest. Third, the multicollinearity problem in the conventional regression analysis can be eliminated in the proposed joint model because the highly correlated variables have been grouped into relatively independent latent factors through the CFA model. Finally, given that the number of latent variables, \( q \), is often much smaller than the number of observed variables, \( p \), the dimension of the risk factors in model (2) can be reduced significantly.

Remark 1. (Model Identification). Model (1) is not identifiable because for any nonsingular matrix \( L \), we have \( Y_i = B \xi_i + \epsilon_i = B L^{-1} \xi_i + \epsilon_i = B^* \xi_i^* + \epsilon_i \), where \( B^* = BL \) and \( \xi_i^* = L^{-1} \xi_i \sim N(0, L^{-1} \Phi(L^{-1})^T) \), indicating that parameters \( B \) and \( \Phi \) are not simultaneously estimable without imposing an identifiability constraint. In this study, we follow a common practice in the literature (Bollen 1989; Lee 2007) to (i) fix the diagonal elements of \( \Phi \) to 1, such that \( \Phi \) is a correlation matrix. This constraint unifies the scale of the latent variables. (ii) Fix additional elements in \( B \) at preassigned values, such that the only possible nonsingular matrix \( L \) that makes \( B^* \) satisfy the constraints is the identity matrix.

Remark 2. In model (1), we make a generic assumption \( E(\xi_i) = 0 \). The extension to the case of \( E(\xi_i) \neq 0 \) is straightforward. When \( E(\xi_i) \neq 0 \), model (2) can be written as \( \lambda(t | Z_i, \xi_i) = \lambda_0(t) + \beta^T Z_i + \gamma^T E(\xi_i) \), and \( \hat{\xi}_i = \hat{\xi}_i - \tilde{E}(\xi_i) \).

Let \( \theta \) be a vector of the unknown parameters in model (1), and \( \hat{\theta} \) be the MLE of \( \theta \). Given that model (1) is a standard CFA model, the estimation of \( \theta \) is relatively simple (Bollen 1989; Shi and Lee 2000; Lee and Song 2004). We first focus on the parameter estimation procedure for model (2), and then outline the EM algorithm to obtain \( \hat{\theta} \) in model (1).

2.2 Borrow-Strength Estimation

We define the counting process \( N_i(t) = I(X_i \leq t, \Delta_i = 1) \) and the at risk process \( Y_i(t) = I(X_i \geq t), i = 1, \ldots, n \). If the
latent variables $\xi$, are observable, $\alpha = (\beta^T, \gamma^T)^T$ can be estimated based on the pseudo-score functions using the approach of Lin and Ying (1994):

$$\hat{U}_1(\alpha) = \sum_{i=1}^{n} \int_{0}^{\tau} \{Z_i - \hat{Z}_i(t)\}[dN_i(t) - Y_i(t)(\beta^T Z_i + \gamma^T \hat{\xi}_i)dt],$$

and

$$\hat{U}_2(\alpha) = \sum_{i=1}^{n} \int_{0}^{\tau} \{\xi_i - \hat{\xi}_i(t)\}[dN_i(t) - Y_i(t)(\beta^T Z_i + \gamma^T \hat{\xi}_i)dt],$$

where $\tau$ is a prespecified positive constant, such that $P(X_i \geq \tau) > 0$, $\hat{Z}_i(t) = S_i^{(1)}(t)/S_i^{(0)}(t)$, $\xi_i = S_i^{(1)}(t)/S_i^{(0)}(t)$, and $S_i^{(0)}(t) = n^{-1}\sum_{i=1}^{n} Y_i(t) Z_i$, $S_i^{(1)}(t) = n^{-1}\sum_{i=1}^{n} Y_i(t) \xi_i$, and $S_i^{(0)} = n^{-1}\sum_{i=1}^{n} Y_i(t)$.

However, the latent variables $\xi$ are unobservable in the proposed joint model. Consequently, the above pseudo-score functions are not directly applicable. Thus, to address this problem, we first estimate $\hat{\xi}_i$ based on model (1). For a given $\theta$, an estimator of $\xi_i$, denoted by $\hat{\xi}_i(\theta)$, can be written as (Lee 2007, chap. 2)

$$\hat{\xi}_i(\theta) = \Gamma(\theta)V_i,$$

where $\Gamma(\theta) = (B^T \Psi_1^{-1} B)^{-1} B^T \Psi_2^{-1}$ is a $q \times p$ matrix function of $\theta$. Although $E[\hat{\xi}_i(\theta)|\xi_i] = \hat{\xi}_i$, a simple replacement of $\xi_i$ by $\hat{\xi}_i(\theta)$ in the estimating functions $U_1(\alpha)$ and $U_2(\alpha)$ would lead to biased estimators because the resulting estimating functions do not have zero mean because $E[\hat{\xi}_i(\theta)|\xi_i] = \hat{\xi}_i$, and $\hat{\xi}_i$ is a prespecified positive constant, such that $\hat{\xi}_i = \Psi_1^{-1} \Psi_2^{-1}$.

In the following M-step have explicit forms

$$E[\hat{\xi}_i(\theta)|\xi_i] = \mu_i + \Psi_2^{-1} \psi_i \theta,$$

where $\mu_i = \psi_i \theta$ and $\psi_i = \psi_1$. The conditional expectations of the sufficient statistics required in the following M-step have explicit forms

$$E[\hat{\xi}_i(\theta)|\xi_i] = \mu_i + \Psi_2^{-1} \psi_i \theta,$$

where $\mu_i = \psi_i \theta$ and $\psi_i = \psi_1$. The conditional expectations of the sufficient statistics required in the following M-step have explicit forms

$$E[\hat{\xi}_i(\theta)|\xi_i] = \mu_i + \Psi_2^{-1} \psi_i \theta.$$



\section{Inference of CFA Model}

Let $V = (V_1, \ldots, V_n)$ and $\xi = (\xi_1, \ldots, \xi_n)$. Given that $\xi$'s are observable, we employed the EM algorithm to obtain $\hat{\theta}$ in model (1) (e.g., Shi and Lee 2000; Lee and Song 2004). The EM algorithm consists of two steps: an E-step for calculating the conditional expectation of the complete-data log-likelihood function with respect to the distribution of $\xi_i$ given the observed data and the current value of $\theta$, as well as an M-step for updating $\theta$ by maximizing the conditional expectation. The complete-data log-likelihood function of model (1) is

$$l(\theta|V, \xi) = \frac{1}{2} \sum_{i=1}^{n} \left( \frac{n + \psi_i \theta}{\Psi_1^{-1} + B^T \Psi_1^{-1} B} \right) \left( \frac{V_i - B \xi_i}{\Psi_1^{-1} + B^T \Psi_1^{-1} B} \right)^T \left( \frac{V_i - B \xi_i}{\Psi_1^{-1} + B^T \Psi_1^{-1} B} \right) + \frac{\theta^T}{\psi_1} \theta,$$

where $\psi_1 = \psi_2$. The $\hat{\theta}$ can be obtained by iterating the E-step and M-step as follows. At the rtth iteration with the current value $\theta^{(r)}$:

\textbf{E-step:} Evaluate $Q(\theta|\theta^{(r)}) = E[l(\theta|V, \xi)|V, \theta^{(r)}] = \int l(\theta|V, \xi)p(\xi|V, \theta^{(r)})d\xi$, where $p(\xi|V, \theta^{(r)}) \equiv N(\mu_i, \Sigma_i)$. The conditional expectations of the sufficient statistics required in the following M-step have explicit forms

$$E[\xi|V, \theta^{(r)}] = \mu_i + \psi_i \theta,$$

where $\mu_i = \psi_i \theta$ and $\psi_i = \psi_1$. The conditional expectations of the sufficient statistics required in the following M-step have explicit forms

$$E[\xi|V, \theta^{(r)}] = \mu_i + \psi_i \theta.$$

\textbf{M-step:} We maximize $Q(\theta|\theta^{(r)})$ with respect to $\theta$ to obtain a new value of $\theta$. Let $B_{ij}$ be the jth row of $B$, $\phi_{ij}$ be the entry in the ith row and jth column of $\Phi$, $\psi_i$ be the jth diagonal element of $\Psi_i$, and $V_{ij}$ be the jth element of $V_i$. The vector of unknown elements in $B_j$ is denoted by $B_{j1}$ and the vector of fixed ones is denoted by $B_{j2}$. The subvectors of $\xi$, corresponding to $B_{j1}$ and $B_{j2}$ are denoted by $\xi_{ij1}$ and $\xi_{ij2}$, respectively. Simple calculations show that the updated estimates of $B_{j1}$ and $\psi_j$ take the following explicit forms

$$B_{j1} = \left[ \sum_{i=1}^{n} E[\xi_{ij2}|V_i, \theta^{(r)}] \right]^{-1} \sum_{i=1}^{n} E[\xi_{ij1}|V_i, \theta^{(r)}],$$

$$\psi_j = n^{-1} \sum_{i=1}^{n} E \left( (V_{ij} - B_{j2}^T \xi_{ij2})^2 | V_i, \theta^{(r)} \right).$$

The updated estimates of unknown parameters in $\Phi$ do not have closed forms, so we use the Newton–Raphson algorithm to calculate them iteratively. Let $\Phi_1$ be a vector of the unknown parameters in $\Phi$.

$$\Phi_1^{new} = \Phi_1^{old} - \left\{ \frac{\partial^2 Q(\theta|\theta^{(r)})}{\partial \Phi_1 \partial \Phi_1^T} \right\}^{-1} \frac{\partial Q(\theta|\theta^{(r)})}{\partial \Phi_1} \bigg|_{\Phi_1^{old}}.$$
The entries in \( \frac{\partial Q(\theta^{(r)})}{\partial \Phi} \) and \( \frac{\partial^2 Q(\theta^{(r)})}{\partial \Phi \partial \Phi_i} \) can be calculated as
\[
\frac{\partial Q(\theta^{(r)})}{\partial \phi_{ij}} = \frac{1}{2} \text{tr} \left( \Phi^{-1} \sum_{i=1}^{n} [E(\xi_i^{\otimes 2}|V_i, \theta^{(r)})] - \Phi \right) \frac{\partial \Phi}{\partial \phi_{ij}}.
\]
\[
\frac{\partial^2 Q(\theta^{(r)})}{\partial \phi_{ij} \partial \phi_{is}} = \frac{1}{2} \text{tr} \left( \Phi^{-1} \frac{\partial \Phi}{\partial \phi_{ij}} \Phi^{-1} \sum_{i=1}^{n} [E(2\xi_i^{\otimes 2}|V_i, \theta^{(r)})] - \Phi \right) \frac{\partial \Phi}{\partial \phi_{is}}.
\]

The expectations required in this step are calculated in the E-step. The convergence of the EM algorithm can be monitored using a commonly used stopping criterion
\[
|l(\theta^{(r)}|V) - l(\theta^{(r-1)}|V)| < \delta,
\]
where \( \delta \) is the predetermined small value (e.g., 0.001) and \( l(\theta|V) \) is the observed-data log-likelihood function
\[
l(\theta|V) = -\frac{1}{2} \left[ np \log(2\pi) + n \log |\Pi(\theta)| + \sum_{i=1}^{n} V_i^T \Pi(\theta)^{-1} V_i \right],
\]
in which \( \Pi(\theta) = B \Phi B^T + \Psi_e \).

The asymptotic covariance matrix of \( \hat{\theta} \) can be calculated by the inverse of the observed information matrix
\[
l(\hat{\theta}) = -\frac{\partial^2 l(\theta|V)}{\partial \theta \partial \theta^T} \bigg|_{\theta=\hat{\theta}}.
\]

Given that the observed information matrix contains complicated matrix differentiation, we approximate it numerically. Let \( l(\theta) = l(\theta|V) \), and let \( \theta_j \) denote the \( j \)th element of \( \theta \), then
\[
\frac{\partial^2 l(\theta)}{\partial \theta_j \partial \theta_k} \bigg|_{\theta=\hat{\theta}} \approx \frac{1}{4m_j m_k} \left[ l(\hat{\theta} + m_j e_j + m_k e_k) + l(\hat{\theta} + m_j e_j - m_k e_k) - l(\hat{\theta} - m_j e_j + m_k e_k) - l(\hat{\theta} - m_j e_j - m_k e_k) \right],
\]
where \( e_j \) is the canonical basis vector that is one at the \( j \)th coordinate, and \( m_j \) is a constant that is small relative to \( \hat{\theta}_j \) (e.g., \( \hat{\theta}_j \times 10^{-3} \)).

**Remark 3.** Shi and Lee (2000) and Lee and Song (2004) also used the EM algorithm to perform the statistical inference for the CFA model. The EM algorithm they employed is actually a Monte Carlo EM (MCEM) algorithm, in which the E-step approximates the conditional expectation via the simulated latent variables and other unknowns via the Gibbs sampling. In the current study, the Monte Carlo step is unnecessary because the conditional expectation can be expressed in a closed form of the parameters. Compared with those in Shi and Lee (2000) and Lee and Song (2004), the inference for the CFA model in the current study is more efficient.

### 3. Asymptotic Properties

Given that the asymptotic properties of \( \hat{\theta} \) have been well studied in the literature (Amemiya, Fuller, and Pantula 1987; Anderson and Amemiya 1988; Lee 2007, pp. 41–43), we analyze the asymptotic properties of \( \hat{\alpha} \), which is of our primary interest in the proposed joint model. Let
\[
K(\theta) = \left[ \Pi(\theta) | \Pi(\theta)^{-1} \otimes \Pi(\theta)^{-1} \right] \hat{\Pi}(\theta)^T \left[ \Pi(\theta) | \Pi(\theta)^{-1} \otimes \Pi(\theta)^{-1} \right]^{-1},
\]
\[
R(\theta) = [\hat{T}_1(\hat{\theta}) K(\theta) \text{vec} \{V_i^{\otimes 2} - \Pi(\theta)\}, \ldots, \hat{T}_p(\theta) K(\theta) \text{vec} \{V_i^{\otimes 2} - \Pi(\theta)\}],
\]
\[
P(\theta) = [\hat{D}_{1i}(\hat{\theta}) K(\theta) \text{vec} \{V_i^{\otimes 2} - \Pi(\theta)\}, \ldots, \hat{D}_{qi}(\theta) K(\theta) \text{vec} \{V_i^{\otimes 2} - \Pi(\theta)\}],
\]
where \( \otimes \) denotes the Kronecker product of two matrices, vec denotes the operation that converts a matrix into a column vector by stacking the rows sequentially, \( \hat{\Pi}(\theta) = \hat{\alpha}(\text{vec} \Pi(\theta))^T / \theta^T \hat{T}_i(\hat{\theta}), i = 1, \ldots, p \) denote the derivatives of the \( i \)th column of \( \Gamma(\theta) \) with respect to \( \theta^T \), and \( \hat{D}_{qi}(\theta) (r = 1, \ldots, q) \) denote the derivatives of the \( r \)th column of \( D(\theta) \) with respect to \( \theta^T \). Define
\[
S_i(t) = n^{-1} \sum_{i=1}^{n} Y_i(t) V_i, \quad \text{and} \quad \tilde{V}(t) = S_i^{(1)}(t) / S^{(0)}(t),
\]
\[
dM_i(t) = dN_i(t) - Y_i(t) [d\hat{\Lambda}_i(t) + \hat{\beta}^T Z_i + \hat{\gamma}^T \hat{T}(\theta) V_i] dt,
\]
\[
\hat{H}_1 = n^{-1} \sum_{i=1}^{n} \int_0^t [Z_i - \tilde{Z}(t)] Y_i(t) V_i^T dt,
\]
\[
\hat{H}_2 = n^{-1} \sum_{i=1}^{n} \int_0^t Y_i(t) [\Gamma(\theta) V_i - \hat{\xi}^*(t; \hat{\theta})] V_i^T dt,
\]
\[
\hat{H}_3 = n^{-1} \sum_{i=1}^{n} \int_0^t [V_i - \tilde{V}(t)] [dN_i(t) - Y_i(t) \hat{\beta}^T Z_i + \hat{\gamma}^T \hat{T}(\theta) V_i] dt,
\]
\[
\hat{H}_4 = n^{-1} \sum_{i=1}^{n} \int_0^t Y_i(t) dt.
\]
Let \( \alpha_0 \) and \( \theta_0 \) be the true values of \( \alpha \) and \( \theta \). Using the consistency of \( \hat{\theta} \), we obtain the consistency of \( \Gamma(\hat{\theta}) \) and \( D(\hat{\theta}) \) (Lemma A.1 of the Appendix), then, we can establish the asymptotic properties of \( \hat{\alpha} \). The results are summarized in the following theorem with the proof given in the Appendix.

**Theorem 1.** Under the regularity conditions (C1)–(C3) as stated in the Appendix, \( \hat{\alpha} \) is consistent to \( \alpha_0 \), and \( n^{1/2} (\hat{\alpha} - \alpha_0) \) has asymptotically a normal distribution with mean zero and covariance matrix that can be consistently estimated by \( \hat{\Sigma}^{-1} \hat{\Sigma} \hat{\Sigma}^{-1} \), where
\[
\Sigma = n^{-1} \sum_{i=1}^{n} \tilde{U}_i^{\otimes 2}, \quad \hat{U}_i = \begin{pmatrix} \hat{U}_{1i} \\ \hat{U}_{2i} \end{pmatrix}, \quad \hat{U}_{1i} = \int_0^t [Z_i - \tilde{Z}(t)] dM_i(t) - \hat{H}_1 R_i(\theta)^T \hat{\gamma},
\]
\[
\hat{U}_{2i} = \int_0^t [\Gamma(\theta) V_i - \hat{\xi}^*(t; \theta)] dM_i(t) + \int_0^t Y_i(t) D(\hat{\theta}) \hat{\gamma} dt - \hat{H}_2 \hat{R}_i(\hat{\theta}) \hat{\gamma} + R_i(\hat{\theta}) \hat{H}_1 + P_i(\hat{\theta}) \hat{\gamma} \hat{H}_4,
\]
\[
\hat{\Sigma} = \begin{pmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{pmatrix}, \quad \hat{A}_{11} = n^{-1} \sum_{i=1}^{n} \int_0^t Y_i(t) [Z_i - \tilde{Z}(t)]^{\otimes 2} dt,
\]
\[
\hat{A}_{12} = n^{-1} \sum_{i=1}^{n} \int_0^t Y_i(t) [Z_i - \tilde{Z}(t)] [\Gamma(\theta) V_i - \hat{\xi}^*(t; \theta)]^T dt,
\]
\[
\hat{A}_{21} = n^{-1} \sum_{i=1}^{n} \int_0^t Y_i(t) [Z_i - \tilde{Z}(t)] \Gamma(\theta) V_i^T dt,
\]
\[
\hat{A}_{22} = n^{-1} \sum_{i=1}^{n} \int_0^t Y_i(t) [Z_i - \tilde{Z}(t)]^2 dt.
\]
Table 1. Results of Simulation 1—AH model, $\xi_j \sim N(0,\Phi)$, $\epsilon_i \sim N(0,0.3I)$

| $n$  | CR | Type (I) $\lambda_0(t)$ | Type (II) $\lambda_0(t)$ | Type (III) $\lambda_0(t)$ |
|------|----|-------------------------|-------------------------|-------------------------|
|     |    | Bias | SE | SEE | CP  | Bias | SE | SEE | CP  | Bias | SE | SEE | CP  |
| 500 | 20%| 0.015 | 0.357 | 0.385 | 0.945 | 0.016 | 0.427 | 0.418 | 0.939 | 0.013 | 0.393 | 0.386 | 0.941 |
|     | 50%| -0.000 | 0.198 | 0.188 | 0.950 | -0.002 | 0.239 | 0.227 | 0.953 | -0.002 | 0.219 | 0.209 | 0.950 |
|     | 1000 | 0.008 | 0.191 | 0.189 | 0.945 | 0.009 | 0.231 | 0.228 | 0.946 | 0.009 | 0.212 | 0.210 | 0.946 |
|     | 20%| 0.009 | 0.433 | 0.442 | 0.951 | 0.006 | 0.489 | 0.499 | 0.954 | 0.006 | 0.452 | 0.458 | 0.953 |
|     | 50%| -0.001 | 0.248 | 0.243 | 0.957 | -0.002 | 0.278 | 0.275 | 0.959 | 0.000 | 0.257 | 0.252 | 0.957 |
|     | 1000 | 0.012 | 0.245 | 0.243 | 0.959 | 0.013 | 0.274 | 0.275 | 0.955 | 0.012 | 0.252 | 0.252 | 0.961 |
|     | 20%| 0.008 | 0.246 | 0.246 | 0.945 | -0.001 | 0.293 | 0.294 | 0.951 | 0.007 | 0.276 | 0.272 | 0.943 |
|     | 50%| -0.002 | 0.128 | 0.132 | 0.966 | 0.001 | 0.162 | 0.160 | 0.957 | 0.004 | 0.156 | 0.148 | 0.942 |
|     | 1000 | 0.005 | 0.134 | 0.133 | 0.946 | 0.001 | 0.164 | 0.160 | 0.945 | 0.004 | 0.149 | 0.148 | 0.954 |
|     | 20%| -0.003 | 0.306 | 0.311 | 0.961 | 0.007 | 0.361 | 0.351 | 0.948 | -0.004 | 0.314 | 0.322 | 0.956 |
|     | 50%| -0.003 | 0.176 | 0.170 | 0.945 | -0.004 | 0.194 | 0.193 | 0.952 | -0.000 | 0.179 | 0.177 | 0.940 |
|     | 1000 | 0.005 | 0.170 | 0.171 | 0.954 | -0.001 | 0.192 | 0.193 | 0.950 | 0.001 | 0.178 | 0.177 | 0.953 |

In model (2), covariate $Z_t$ was independently drawn from the Bernoulli distribution with success probability of 0.5. The true population values of the parameters were set as $\beta = 1$ and $\gamma = (0.2,0.5)^T$. The failure time $T_i$ was generated based on model (2) with three types of the baseline hazard function: (I) $\lambda_0(t) = 3$ (constant), (II) $\lambda_0(t) = 4t + 3$ (linear), and (III) $\lambda_0(t) = 6t^2 + 3$ (nonlinear). The censoring time $C_i$ was generated from a uniform distribution on $(0, c)$, where $c$ was selected as $\{1.5, 1.2, 1.3\}$ and $\{0.47, 0.40, 0.45\}$ to yield two levels of censoring rate (CR), 20% and 50%, for the three types of $\lambda_0(t)$.

In implementing the EM algorithm to estimate $\hat{\theta}$, the convergence was monitored and claimed with $\delta = 0.001$ (see (4)). The $\hat{\alpha}$ was then obtained using (3). The results presented below are based on 1000 replications with sample sizes $n = 500$ and 1000.

Table 2. Results of Simulation 1—CFA model: $\xi_j \sim N(0,\Phi)$, $\epsilon_i \sim N(0,0.3I)$

| $n$ | CR | Para | Bias | SE | SEE | CP |
|-----|----|------|------|---|---|---|
| 500 | 20% | $b_{11}$ | 0.001 | 0.039 | 0.038 | 0.944 |
|     | 50% | $b_{21}$ | 0.001 | 0.039 | 0.038 | 0.948 |
|     | 1000 | $b_{31}$ | 0.000 | 0.038 | 0.038 | 0.954 |
|     | 20% | $b_{41}$ | 0.001 | 0.038 | 0.038 | 0.952 |
|     | 50% | $b_{51}$ | 0.000 | 0.037 | 0.038 | 0.956 |
|     | 1000 | $b_{61}$ | 0.000 | 0.050 | 0.050 | 0.945 |
|     | 20% | $\phi_{12}$ | -0.001 | 0.031 | 0.030 | 0.944 |
|     | 50% | $\phi_{22}$ | -0.001 | 0.030 | 0.030 | 0.949 |
|     | 1000 | $\phi_{32}$ | -0.001 | 0.030 | 0.030 | 0.941 |
|     | 20% | $\psi_{1}$ | -0.001 | 0.031 | 0.030 | 0.944 |
|     | 50% | $\psi_{2}$ | -0.001 | 0.030 | 0.030 | 0.949 |
|     | 1000 | $\psi_{3}$ | -0.001 | 0.030 | 0.030 | 0.941 |
|     | 20% | $\psi_{4}$ | -0.001 | 0.030 | 0.030 | 0.950 |
|     | 50% | $\psi_{5}$ | -0.001 | 0.030 | 0.030 | 0.942 |
|     | 1000 | $\psi_{6}$ | -0.002 | 0.029 | 0.030 | 0.956 |
The main goal of this simulation is to investigate the sensitivity of the proposed method to the normality assumption of the latent variables and the random errors. As suggested by one of the anonymous reviewers, we consider several nonnormal cases, including the severely skewed distributions used in Huang and Wang (2000). The model setup is the same as that of Simulation 1 except Case (1): $\xi_1 \sim N(\mu_1, \Sigma_1) + N(\mu_2, \Sigma_2)$, where $\mu_1 = (-0.5, -0.5)^T$, $\mu_2 = (1.0, 1.0)^T$, and $\{\sigma_{11}, \sigma_{12}, \sigma_{22}\}$ in $\Sigma_1 = \Sigma_2$ are $(0.5, -0.3, 0.5)$; Case (2): $\xi_j \sim \{\text{Gamma}(4, 2) - 2\}|I_2$, where Gamma($\alpha_1, \alpha_2$) stands for the gamma distribution with shape parameter $\alpha_1$ and rate parameter $\alpha_2$, and $I_2$ is a two-dimensional identity matrix; Case (3): $\xi_j \sim \text{beta}(3, 1)|I_2$, where beta($\alpha_1, \alpha_2$) stands for the beta distribution with shape parameters $\alpha_1$ and $\alpha_2$; Case (4): $\epsilon_{ij} \sim \text{Gamma}(5, 4) - 5/4$ for $j = 1, \ldots, 6$, where $\epsilon_{ij}$ is the $j$th element of $\epsilon_i$; Case (5): $\epsilon_{ij} \sim \xi_3$, where $\xi_3$ stands for the $t$ distribution with degrees of freedom 3; and Case (6): $\epsilon_{ij} \sim \text{beta}(3, 1)$.

The results with Type (I) $\lambda_{0}(t)$ and $n = 500$ under Cases (1)–(6) are presented in Tables S4 and S5 of supplementary material. Most of the parameter estimates, especially those related to the AH model, are quite robust to the nonnormality of $\xi_j$ and $\epsilon_i$. While nonnormal $\xi_j$ mainly influence the standard error estimates of the factor loadings in $B$, nonnormal $\epsilon_i$ mostly impact on those of the error variances in $\Psi$. We also checked the sensitivity of the estimation results when $\xi_j$ and $\epsilon_i$ follow other nonnormal distributions. Similarly, the inference related to the AH model is very robust to the nonnormality of $\xi_j$ and $\epsilon_i$, whereas the standard errors of some parameters in the CFA model are overestimated/underestimated. Considering that the primary goal of the current study is to investigate the potential risk factors of hazards of interest, the inaccurate standard error estimates of error variances in the CFA model would not cause a problematic inference of the AH model.

5. APPLICATION TO CKD DATA

The proposed methodology was applied to a study of CKD for Type 2 diabetic patients as discussed in Section 1. The main objective is to investigate the potential risk factors that affect the hazards of CKD. The failure time of CKD was calculated as the periods from enrollment to the date of the first clinical endpoint or January 31, 2009, whichever came first. The failure (clinical endpoint) of CKD was defined by diabetic nephropathy (DNP) plus follow-up estimate glomerular filtration rate (eGFR) (clinical endpoint) of CKD was defined by diabetic nephropathy (DNP) plus follow-up estimate glomerular filtration rate (eGFR). The data generated from this new model, we conduct the analysis with a misspecified $B$, wherein nonzero $b_{81}$ is misspecified as zero. Table S2 of supplementary material shows that the estimation results related to the AH model are quite robust, but those related to the CFA model are slightly influenced under the case considered. Nevertheless, it is worthwhile to note that the statistical inference for the CFA is sensitive to misspecification of the number of latent factors. When fitting a wrong model with extra (nonexistent) or missing (existent) latent factors, we found that the computer algorithm always diverged.

Further, we examined the empirical performance of the proposed method in which the error variance increases. We again used the previous simulation setting but increased the error variance $\psi_{1j}, j = 1, \ldots, 6$ from 0.3 to 1.0. The results obtained with Type (1) $\lambda_{0}(t)$ and $n = 500$ are reported in Table S3 of supplementary material. Except for slightly increased standard error estimates, the parameter estimates are still unbiased and close to the ones obtained under smaller error variance.

### Table 3. The factor loadings obtained based on EFA through LISREL

| Variable | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Unique variance |
|----------|----------|----------|----------|----------|-----------------|
| WAIST    | 1.005    | 0.009    | -0.057   | -0.025   | 0.000           |
| BMI      | 0.830    | 0.021    | 0.029    | -0.023   | 0.307           |
| SBP      | 0.005    | 0.839    | 0.038    | -0.039   | 0.311           |
| DBP      | 0.024    | 0.737    | 0.007    | 0.001    | 0.449           |
| HbA1c    | 0.040    | -0.040   | -0.963   | -0.047   | 0.102           |
| FPG      | 0.000    | 0.012    | -0.723   | 0.019    | 0.470           |
| TC       | -0.072   | 0.151    | -0.158   | 0.303    | 0.825           |
| HDL-C    | -0.245   | 0.052    | -0.063   | -0.271   | 0.843           |
| TG       | -0.033   | -0.034   | 0.040    | 1.019    | 0.000           |
Figure 1. Path diagram of the proposed joint model, along with parameter estimates and their standard error estimates (in parentheses), in the analysis of the CKD data. In the path diagram, the latent variables are enclosed by ellipses, whereas the observed variables are enclosed by squares.

Let $\mathbf{V} = (V_1, V_2, \ldots, V_9)^T = (\text{WAIST, BMI, SBP, DBP, HbA1c, FPG, TC, HDL-C, TG})^T$, each of which is standardized prior to the analysis. $\xi = (\xi_1, \xi_2, \xi_3, \xi_4)^T$ = (obesity, blood pressure, glycemia, lipid)$^T$, and $\mathbf{Z} = (Z_1, Z_2, Z_3)^T$ = (age at enrollment, duration of diabetes, sex)$^T$. We employed models (1) and (2) with $p = 9$, $q = 4$, $s = 3$, and the following nonoverlapping factor loading matrix to conduct the analysis:

$$
B^T = \begin{bmatrix}
b_{11} & b_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
b_{32} & b_{42} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
b_{53} & b_{63} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
b_{74} & b_{84} & b_{94}
\end{bmatrix},
$$

where the zeros are fixed elements.

In estimating $\hat{\theta}$, we set $\delta = 0.001$. The $\hat{\theta}$ and $\hat{\alpha}$ were then obtained based on the EM algorithm and (3), and their standard error estimates were obtained based on (5), (6), and Theorem 1.

The proposed joint model and the analytical results are reported in Figure 1. For clarity, the less important parameters in $\Psi_\epsilon$ and $\Phi$ are not reported. Several findings are obtained from Figure 1. First, the age at enrollment has significantly positive effect on the hazards of CKD, indicating that older patients have higher risks of developing CKD. Second, the duration of diabetes also has significantly positive effect on the hazards of CKD. Patients with longer duration of diabetes are more predisposed to CKD. Third, different from age and duration of diabetes, the effect of sex on CKD is not significant. Thus, male and female diabetic patients have equal risks of suffering this complication. Fourth, all latent risk factors have
patients and to improve their quality of life. Meticulous monitoring and an early diagnosis of diabetes are essential to prevent CKD or other complications for Type 2 diabetic patients. These findings have public health implications, especially for the aggressive control of risk factors that are confounding and hard to interpret. By contrast, correlations elicit multicollinearity and result in diverse effects of the observed indicators of a latent risk factor. For instance, the effect on the hazards of CKD is significant for FPG, and negative for TC, nonsignificant for HDL-C, significant for HbA1c but nonsignificant for WAI ST but nonsignificant for BMI, positive for Lipid and provides important insights for the latent risk factors of diabetic kidney disease. Important insights for preventing such complication were obtained.

For comparison, we conducted a standard analysis by regarding the observed variables, $V_1, \ldots, V_9$, as independent covariates in the AH model. The results are reported in Table 4, which shows the diverse effects of the observed indicators of a latent risk factor. For instance, the effect on the hazards of CKD is significant for WAI ST but nonsignificant for BMI, positive for SBP but negative for DBP, significant for HbA1c but nonsignificant for FPG, and negative for TC, nonsignificant for HDL-C, but positive for TG. A further check reveals high correlations among the observed variables. For instance, the sample correlation between WAI ST and BMI, SBP and DBP, and HbA1c and FPG, are 0.827, 0.615, and 0.687, respectively. These high correlations elicit multicollinearity and result in diverse effects that are confounding and hard to interpret. By contrast, the proposed joint analysis avoids the multicollinearity problem and provides important insights for the latent risk factors of CKD.

Given that there is a slight disagreement between the LISA REL analysis ($b_{31} = -0.245$) and our specification ($b_{31} = 0$) for the CFA model, we reanalyzed the dataset under the same setup except allowing $b_{31} \neq 0$. The parameter estimates, along with those obtained under $b_{31} = 0$, are reported in Table S6 of supplementary material. The two analyses yield very similar results. In particular, the parameter estimates related to the AH model are identical up to the first three decimals. A sensitivity analysis in Simulation 1 also demonstrated the robustness of the proposed method to such a misspecification. Although we adopted a nonoverlapping structure of $B$ based on the suggestion of medical experts and the existing medical literature in this study, the use of overlapping or nonoverlapping $B$ in substantive research should be decided on a problem-to-problem basis.

6. DISCUSSION

In this article, we propose a joint model to investigate how observed and latent risk factors influence the hazard rate of interest. The proposed model is novel and advantageous over conventional AH models in terms of both analytic power and interpretability. A hybrid procedure that combines the EM algorithm and a borrow-strength estimation method is proposed to conduct the statistical inference. The consistency and asymptotic normality of the parameter estimators are established. Simulation results demonstrate that the proposed method performed well. The novel model and methodology were applied to a real problem of identifying potential risk factors of diabetic kidney disease. Important insights for preventing such complication were obtained.

The borrow-strength estimation procedure developed in this article is essentially a two-stage approach. The first stage estimates the latent variables and the parameters of the CFA model by the EM algorithm, whereas the second stage estimates the regression coefficients of the AH model by using the estimated latent variables with corrected pseudo-score functions. A possible alternative estimation method is to implement the EM algorithm based on the complete-data likelihood function of the proposed joint model. However, this complete-data likelihood function involves high-dimensional integration and unspecified baseline hazard function. The nonlinearity of the AH model makes the E-step extremely challenging because $f(\xi_i|V_i, X_i, \Delta_i, \theta, \alpha, \lambda_0(t))$ is very complicated, and thus sampling becomes difficult. In addition, the unspecified baseline hazard function leads to a tedious maximization in the M-step. Thus, the feasibility of such complete-data likelihood-based method is uncertain and requires further investigation.

Our approach can be regarded as a generalization of the conventional AH model to a broader class of models through the incorporation of latent variables. To the best of our knowledge, our study is the first to introduce latent variables into the survival modeling framework. Further, our approach can be extended to more generalizable models. First, the current study requires the normal assumption of the latent variables and the measurement error. However, officially assessing this assumption is
impossible because the latent variables are not directly observable. Although the previous simulation demonstrates that the inference of the AH model is robust to the violation of this normal assumption, it is of interest to develop novel methods that do not require such assumption. One promising direction in relaxing this assumption is considering an asymptotically distribution-free approach (Brownie 1984) to obtain the parameter estimation of the CFA model and then modifying the corrected pseudo-score method accordingly to perform the statistical inference for the AH model. Second, our model can also be extended to a longitudinal setting, in which the CFA model is applicable to the additive-multiplicative hazards model that is of interest by the associate editor, considering linear transformation of \(\tilde{\theta}\), \(\tilde{\beta}\), \(\tilde{\gamma}\), and \(\tilde{\xi}\), are defined similarly as in the CFA-AH joint model. A standard analysis of the PH model is based on estimating function \(\hat{U}_p(\alpha) = (\hat{U}_{p1}(\alpha)^T, \hat{U}_{p2}(\alpha)^T)^T\) with \(\hat{U}_{p1}(\alpha)\) and \(\hat{U}_{p2}(\alpha)\) defined similarly as in the CFA-AH joint model. Here, the correction of \(\hat{U}_p(\alpha)\) pertaining to exponential functions of \(\tilde{\xi}\) is more difficult than that corresponding to quadratic functions of \(\tilde{\xi}\). Inspired by the previous work (e.g., Nakamura 1992; Kong and Gu 1999) on PH model with covariate measurement error, we propose corrected estimating function \(U_p(\alpha; \theta) = (U_{p1}(\alpha; \theta)^T, U_{p2}(\alpha; \theta)^T)^T\) with

\[
U_{p1}(\alpha; \theta) = \sum_{i=1}^{n} \int_{t_0}^{t} \left[ Z_i - \sum_{j=1}^{n} Y_j(t) \exp(\tilde{\beta}^T Z_j + \tilde{\gamma}^T \tilde{\xi}_j) Z_j \right] dN_i(t),
\]

\[
U_{p2}(\alpha; \theta) = \sum_{i=1}^{n} \int_{t_0}^{t} \left[ \tilde{\xi}_i - \sum_{j=1}^{n} Y_j(t) \exp(\tilde{\beta}^T Z_j + \tilde{\gamma}^T \tilde{\xi}_j) \tilde{\xi}_j \right] dN_i(t).
\]

The estimator of \((\alpha, \theta)\) can be obtained similarly. As expected, the derivation of asymptotic properties of the parameter estimates for CFA-PH joint model will be more technically involved. The details are not reported here. The same rationale is applicable to the additive-multiplicative hazards model that subsumes both PH and AH models as special cases. As suggested by the associate editor, considering linear transformation models (Zeng, Yin, and Ibrahim 2005; Zeng, Lin, and Lin 2008) is a promising attempt for formulating the approach in a more general way, but it requires new theoretical and computational developments. The feasibility of obtaining a corrected estimating function analytically under the transformation model framework requires further investigation.

**APPENDIX: PROOFS OF ASYMPTOTIC RESULTS**

We use the same notation defined earlier and take all limits at \(n \to \infty\). Let \(s^{11}(t), s^{12}(t),\) and \(s^{00}(t)\) be the limits of \(S^{11}(t)/\Gamma_1(t), S^{12}(t)/\Lambda_1(t),\) and \(S^{00}(t)/\Gamma_2(t),\) respectively. To study the asymptotic properties of the proposed estimators, we need the following regularity conditions:

(C1) The matrix \(\Pi(\theta_0)\) is positive definite; all partial derivatives of the first three orders of \(\Pi(\theta)\) with respect to the elements of \(\theta\) are continuous and bounded in a neighborhood of \(\theta_0\); the matrix \(\Pi(\theta)\) is of full rank in a neighborhood of \(\theta_0\).

(C2) \(A_0(t) < \infty, P(X_i \geq t) > 0,\) and \(Z_i\) is bounded almost surely.

(C3) \(A\) is nonsingular, and

\[
A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix}, \quad A_{11} = E \left[ \int_0^t Y_j(t) |Z_i - e_i(t)|^{p/2} dt \right],
\]

\[
A_{12} = E \left[ \int_0^t Y_j(t) |Z_i - e_i(t)| \tilde{\xi}_j dt \right],
\]

\[
A_{22} = E \left[ \int_0^t Y_j(t) \tilde{\xi}_j^{p/2} dt \right],
\]

where \(e_i(t) = s^{11}(t)/s^{00}(t),\) and \(\tilde{\xi}_i(t) = s^{11}(t)/s^{00}(t).

Note that under model (1), \(V_i\) is distributed as \(N(0, \Pi(\theta))\). Let \(S^* = n^{-1} \sum_{i=1}^{n} V_i^{p/2}\). The likelihood function of \(V = (V_1, \ldots, V_n)\) is

\[
L(\theta|V) = (2\pi)^{-n^{p/2}} |\Pi(\theta)|^{-n/2} \exp \left[ -\frac{1}{2} \sum_{i=1}^{n} V_i^T \Pi(\theta)^{-1} V_i \right],
\]

and the \(\hat{\theta}\) is the maximizer of \(L(\theta|V)\). Some straightforward calculation yields that \(\hat{\theta}\) is the minimizer of the following discrepancy function:

\[
F(\theta) = \log |\Pi(\theta)| + tr(S^* \Pi(\theta)^{-1}) + p \log(2\pi).
\]

To prove Theorems 1 and 2, we need the following lemma.

**Lemma A.1.** Under the condition (C1), \(\Gamma(\hat{\theta})\) and \(D(\hat{\theta})\) are consistent to \(\Gamma(\theta_0)\) and \(D(\theta_0)\), respectively. Furthermore,

\[
\hat{\theta} - \theta_0 = K(\theta_0)\text{vec}(S^* - \Pi(\theta_0)) + o_p(n^{-1/2}),
\]

\[
\Gamma(\hat{\theta}) - \Gamma(\theta_0) = n^{-1} \sum_{i=1}^{n} R_i(\theta_0) + o_p(n^{-1/2}),
\]

\[
D(\hat{\theta}) - D(\theta_0) = n^{-1} \sum_{i=1}^{n} P_i(\theta_0) + o_p(n^{-1/2}),
\]

where \(K(\theta), R_i(\theta),\) and \(P_i(\theta)\) are defined in the first paragraph of Section 3.

**Proof.** The maximum likelihood estimation theory of the CFA model (Amemiya, Fuller, and Pantula 1987; Anderson and Amemiya 1988; Lee 2007, pp. 41–43) indicates that \(\hat{\theta}\) is consistent to \(\theta_0\), that is, \(\hat{\theta}\) converges in probability to \(\theta_0\). In addition, using the continuous mapping theorem and consistency of \(\hat{\theta}\), we can obtain the consistency of \(\Gamma(\hat{\theta})\) and \(D(\hat{\theta})\). Let \(F^*(\theta) = \frac{\partial F(\theta)}{\partial \theta}\), and \(\hat{F}^*(\theta) = \frac{\partial F(\theta)}{\partial \hat{\theta}}\). A straightforward calculation shows that

\[
\hat{F}^*(\theta) = -\Pi(\theta)(\Pi(\theta)^{-1} \otimes \Pi(\theta)^{-1})\text{vec}(S^* - \Pi(\theta)),
\]
and
\[ F^*(\hat{\theta}) = \hat{\Pi}(\hat{\theta})\hat{\Pi}(\hat{\theta})^{-1} \hat{\Pi}(\hat{\theta})^T - 2\hat{\Pi}(\hat{\theta})\hat{\Pi}(\hat{\theta})^{-1}(S' - \hat{\Pi}(\hat{\theta}))\hat{\Pi}(\hat{\theta})^T - \hat{\Pi}(\hat{\theta})(I_p \otimes (\hat{\Pi}(\hat{\theta})^{-1} \hat{\Pi}(\hat{\theta})^{-1})\text{vec}(S' - \hat{\Pi}(\hat{\theta}))), \]
where \( \rho \) is the dimension of \( \theta \). The uniform law of large numbers implies that \( F^*(\hat{\theta}) \) converges in probability to a nonrandom matrix function uniformly in \( \theta \). In particular, \( F^*(\hat{\theta}) \) converges in probability to \( \hat{\Pi}(\hat{\theta})\hat{\Pi}(\hat{\theta})^{-1} \hat{\Pi}(\hat{\theta})^T \) for \( k = 1, \ldots, \rho \), let \( F^*_k(\hat{\theta}) \) denote the \( k \)-th element of \( F^*(\hat{\theta}) \). It follows from the Taylor expansion that
\[ F^*_k(\hat{\theta}) = \hat{F}^*_k(\hat{\theta}_0) + \frac{\partial \hat{F}^*_k(\hat{\theta})}{\partial \theta} |_{\hat{\theta} = \hat{\theta}_0}(\hat{\theta} - \theta_0). \]
where \( \theta^k \) is on the line segment between \( \hat{\theta} \) and \( \theta_0 \). Note that \( F^*(\hat{\theta}) = 0 \).

By the consistency of \( \hat{\theta} \), one can show that \( \theta^k \) converges in probability to \( \theta_0 \). Then, \n\[
\left( \frac{\partial \hat{F}^*_k(\hat{\theta})}{\partial \theta}, \ldots, \frac{\partial \hat{F}^*_k(\hat{\theta})}{\partial \theta} \right)_{| \theta_0 = \theta_0}^T
\]
converges in probability to the limit of \( \hat{F}^*(\hat{\theta}_0) \), which is nonsingular by the regularity conditions. Note that by the multivariate central limit theorem, \( \hat{F}^*(\hat{\theta}_0) = O_p(n^{-1/2}). \) Using the uniform convergence of \( \hat{F}^*(\hat{\theta}) \) and consistency of \( \theta \), we obtain (A.1). Similarly, by the Taylor expansion,
\[ \Gamma(\hat{\theta}) - \Gamma(\theta_0) = \left[ \hat{\Gamma}_1(\theta_0)(\hat{\theta} - \theta_0), \ldots, \hat{\Gamma}_\rho(\theta_0)(\hat{\theta} - \theta_0) \right] + o_p(n^{-1/2}), \]
which together with (A.1) gives (A.2). In a similar manner, we obtain (A.3).

**Proof of Theorem 1.** Define
\[ dM_i(t) = dN_i(t) - Y_i(t)\lambda_i(t) + \beta_i^T Z_i + \gamma_i^T \Gamma(\theta_0) V_i dt. \]

Note that \( E \{ \Gamma(\theta_0) V_i | \xi_i \} = \xi_i \). Then, under model (2), \( M_i(t) \) are zero-mean stochastic processes, \( i = 1, \ldots, n \). Thus, it follows from Lemma A.1 of Lin and Yang (2001) that
\[ U_1(\alpha_0; \theta_0) = \sum_{i=1}^n \int_0^T (Z_i - \bar{Z}(t))dM_i(t) = \sum_{i=1}^n \int_0^T (Z_i - \bar{Z}(t))dM_i(t) + o_p(n^{1/2}). \] (A.4)

Using (A.2), we have
\[ U_2(\alpha_0; \hat{\theta}) - U_2(\alpha_0; \theta_0) = \hat{H}_0(\Gamma(\theta_0) - \Gamma(\theta_0))^T y_0 = - \sum_{i=1}^n H_i R_i(\theta_0)^T y_0 + o_p(n^{1/2}), \] (A.5)
where \( \hat{H}_0 \) is defined in Section 3, and \( H_i \) is the limit of \( \hat{H}_i \). It follows from (A.4) and (A.5) that
\[ U_1(\alpha_0; \hat{\theta}) = U_1(\alpha_0; \theta_0) + U_1(\alpha_0; \hat{\theta}) - U_1(\alpha_0; \theta_0) = \sum_{i=1}^n U_i + o_p(n^{1/2}), \] (A.6)

where
\[ \int_0^T (Z_i - \bar{Z}(t))dM_i(t) = H_i R_i(\theta_0)^T y_0. \]

Following arguments similar to those in the proof of (A.4), we obtain
\[ U_2(\alpha_0; \theta_0) = \sum_{i=1}^n \int_0^T (\xi_i(\theta_0) - \xi_i(t; \theta_0))dM_i(t) + \sum_{i=1}^n \int_0^T Y_i(t) D(\theta_0) y_{i\theta} dt \]
\[ = n \sum_{i=1}^n \int_0^T (\Gamma(\theta_0) V_i - e_i(t))dM_i(t) + \sum_{i=1}^n \int_0^T Y_i(t) D(\theta_0) y_{i\theta} dt + o_p(n^{1/2}). \] (A.7)

Some straightforward calculation yields
\[ U_2(\alpha_0; \hat{\theta}) - U_2(\alpha_0; \theta_0) = -n\hat{H}_0^T(\Gamma(\hat{\theta}) - \Gamma(\theta_0))^T y_0 + n(\Gamma(\hat{\theta}) - \Gamma(\theta_0))^T \hat{H}_0^T \]
\[ + n(\Gamma(\hat{\theta}) - D(\theta_0))^T y_0 dt. \]

and \( \hat{H}_0 \) is defined in Section 3. Using Lemma 1, we obtain
\[ U_2(\alpha_0; \hat{\theta}) = U_2(\alpha_0; \theta_0) + \sum_{i=1}^n H_i R_i(\theta_0)^T y_0 + o_p(n^{1/2}) \]
\[ = \sum_{i=1}^n U_i + o_p(n^{1/2}), \] (A.8)

where \( H_i, H_0, \) and \( H_0 \) are the limits of \( \hat{H}_i, \hat{H}_0, \) and \( H_0 \), respectively. Thus, using (A.7) and (A.8), we obtain
\[ U_2(\alpha_0; \hat{\theta}) = U_2(\alpha_0; \theta_0) + \sum_{i=1}^n (U_2(\alpha_0; \hat{\theta}) - U_2(\alpha_0; \theta_0)) \]
\[ = \sum_{i=1}^n U_i + o_p(n^{1/2}), \] (A.9)

Let \( U_i = (U_{i1}^T, U_{i2}^T)^T \). Then it follows from (A.6) and (A.9) that
\[ U(\alpha_0; \hat{\theta}) = \sum_{i=1}^n U_i + o_p(n^{1/2}). \] (A.10)

which is a sum of \( n \) independent identically distributed zero-mean random vectors plus an asymptotically negligible term. The law of large numbers and the multivariate central limit theorem show that \( n^{-1}U(\alpha_0; \hat{\theta}) \rightarrow 0 \) in probability and \( n^{-1/2}U(\alpha_0; \hat{\theta}) \) converges in distribution to a normal random vector with mean zero and covariance matrix \( \Sigma = E[U_i^2] \). Note that
\[ \hat{a} - \alpha_0 = n^{-1}\hat{A}^{-1}U(\alpha_0; \hat{\theta}), \] (A.11)
and \( \hat{A} \rightarrow \Lambda \) in probability by the consistency of \( \Gamma(\hat{\theta}) \) and \( D(\hat{\theta}) \). Then, based on (A.11), \( \hat{a} \) converges in probability to \( \alpha_0 \), and \( n^{1/2}(\hat{a} - \alpha_0) \) is asymptotically normal with mean zero and covariance matrix \( \Lambda^{-1} \Sigma \Lambda^{-1} \), which can be consistently estimated by \( \hat{\Lambda}^{-1} \hat{\Sigma} \hat{\Lambda}^{-1} \). \( \square \)
Proof of Theorem 2. Note that
\[
d N_t(u) - Y_t(u) \left( \hat{\beta}^T Z_t + \hat{v}^T \Gamma(\hat{\theta}) V_t \right) du - Y_t(u) \hat{\lambda}_0(u) du
\]
\[
= d M_t(u) - Y_t(u) \left( \hat{\beta} - \beta_0 \right)^T Z_t + \hat{v}^T \Gamma(\hat{\theta}) - \Gamma(\theta_0) \right) V_t du + (\hat{\gamma} - \gamma_0)^T \Gamma(\hat{\theta}_0) V_t du.
\]
Thus,
\[
\hat{\lambda}_0(t) - \lambda_0(t) = n^{-1} \int_0^t \sum_{i=1}^n O_i(t) + o(n^{-1/2}),
\]
(A.12)
where
\[
O_i(t) = \int_0^t d M_t(u) - s^{(0)}(u) - \gamma_i^T R_i(\theta_0) \int_0^t e_i(u) du - e_i(t)^T A^{-1} U_i.
\]
In view of the consistency of \( \hat{\alpha} \), from the uniform law of large numbers (Pollard 1990) and the multivariate central limit theorem, \( \hat{\alpha}(t) \) converges in probability to \( \hat{\lambda}_0(t) \) uniformly in \( t \in [0, \tau] \), and \( n^{1/2}(\hat{\alpha}(t) - \alpha(t)) \) converges in finite-dimensional distributions to a zero-mean Gaussian process. Since \( O_i(t) (i = 1, \ldots, n) \) can be written as sums or products of monotone functions of \( t \) and are thus tight (van der Vaart and Wellner 1996). Thus, \( n^{1/2}(\hat{\alpha}(t) - \alpha(t)) \) is tight and converges weakly to a zero-mean Gaussian process with covariance function at \( (t, s) \) given by \( \Sigma(t, s) = E(O_i(t) O_i(s)) \), which can be consistently estimated by \( \hat{\Sigma}(t, s) \) defined in Theorem 2. □

SUPPLEMENTARY MATERIALS

Tables S1–S6 are presented in the online supplementary materials.

[Received October 2013. Revised May 2014]

REFERENCES

Aalen, O. O. (1989), “A Linear Regression Model for the Analysis of Life Times,” Statistics in Medicine, 8, 907–925. [1148]

Anemiy, Y., Fuller, W. A., and Pantula, S. G. (1987), “The Asymptotic Distribution of Some Estimators for a Factor Analysis Model,” Journal of Multivariate Analysis, 22, 51–64. [1152,1157]

Anderson, T. W., and Amemiya, Y. (1988), “The Asymptotic Normal Distribution of Estimators in Factor Analysis Under General Conditions,” The Annals of Statistics, 16, 759–771. [1152,1157]

Bentler, P. M. (2002), EQS: Structural Equations Program Manual, Encino, CA: Multivariate Software. [1149]

Bollen, K. A. (1989), Structural Equations with Latent Variables, New York: Wiley. [1148,1150]

Breslow, N. E., and Day, N. E. (1987), Statistical Methods in Cancer Research, Vol. II, The Design and Analysis of Cohort Studies, Lyon: IARC. [1148]

Brownie, M. W. (1984), Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures,” British Journal of Mathematical and Statistical Psychology, 37, 62–83. [1157]

Cai, J., and Zeng, D. (2011), “Additive Mixed Effect Model for Clustered Failure Time Data,” Biometrics, 67, 1340–1351. [1150]

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995), Measurement Error in Nonlinear Models, London: Chapman and Hall. [1149]

Cox, D. R. (1972), “Regression Models and Life-Tables” (with discussion), Journal of the Royal Statistical Society, Series B, 34, 187–220. [1148]

Cox, D. R., and Oakes, D. (1984), Analysis of Survival Data, London: Chapman and Hall. [1148]

Huang, Y., and Wang, C. Y. (2000), “Cox Regression With Accurate Covariates Uncertainable: A Nonparametric-Correction Approach,” Journal of the American Statistical Association, 95, 1209–1219. [1154]

Huffer, F. W., and McKeague, I. W. (1991), “Weighted Least Squares Estimation for Aalen’s Additive Risk Model,” Journal of the American Statistical Association, 86, 114–129. [1148]

Jøreskog, K. G., and Sörbom, D. (1996), LISREL 8: Structural Equation Modeling With the SIMPLIS Command Language, Lincolnwood, IL: Scientific Software International. [1149,1149]

Kong, F. H., and Gu, M. (1999), “Consistent Estimation in Cox Proportional Hazards Model With Covariate Measurement Errors,” Statistica Sinica, 9, 953–969. [1157]

Kulich, M., and Lin, D. Y. (2000), “Additive Hazards Regression With Covariate Measurement Error,” Journal of the American Statistical Association, 95, 238–248. [1149]

Lawley, D. N., and Maxwell, A. E. (1971), Factor Analysis as a Statistical Method, London: Butterworths. [1148]

Lee, S. Y. (2007), Structural Equation Modeling: A Bayesian Approach, New York: Wiley. [1148,1150,1151,1152,1157]

Lee, S. Y., and Song, Y. X. (2004), “Maximum Likelihood Analysis of a General Latent Variable Model With Hierarchically Mixed Data,” Biometrics, 60, 624–636. [1150,1151,1152]

Lin, D. Y., and Ying, Z. (1994), “Semiparametric Analysis of the Additive Risk Model,” Biometrika, 81, 61–71. [1148,1151]

Muthén, L. K., and Muthén, B. O. (1998–2007), Mplus User’s Guide (5th ed.), Los Angeles, CA: Muthén & Muthén. [1149]

Nakamura, T. (1992), “Proportional Hazards Model With Covariates Subject to Measurement Error,” Biometrics, 48, 829–838. [1157]

Pollard, D. (1990), Empirical Processes: Theory and Applications, Hayward, CA: IMS. [1159]

Roy, J., and Lin, X. (2000), “Latent Variable Models for Longitudinal Data With Multiple Continuous Outcomes,” Biometrics, 56, 1047–1054. [1149]

Sammel, M. D., and Ryan, L. M. (1996), “Latent Variable Models With Fixed Effects,” Biometrics, 52, 650–663. [1149]

Shi, J. Q., and Lee, S. Y. (2000), “Latent Variable Models With Mixed Continuous and Polytomous Data,” Journal of the Royal Statistical Society, Series B, 62, 77–87. [1149,1150,1151,1152]

Song, Y., and Huang, Y. (2006), “A Corrected Pseudo-Score Approach for Additive Hazards Model With Longitudinal Covariates Measured With Error,” Lifetime Data Analysis, 12, 97–110. [1149,1150]

Song, X. Y., Lee, S. Y., Ma, R. C. W., So, W. Y., Cai, J. H., Tam, C., Lam, V., Ying, W., Ng, M. C. Y., and Chan, J. C. N. (2009), “Phenotype-Genotype Interactions on Renal Function in Type 2 Diabetes: An Analysis Using Structural Equation Modelling," Diabetologia, 52, 1543–1553. [1149,1154]

Song, X. Y., and Lee, S. Y. (2012), Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences, London: Wiley. [1148,1150]

van der Vaart, A. W., and Wellner, J. A. (1996), Weak Convergence and Empirical Processes, New York: Springer-Verlag. [1159]

Wang, Y., Cai, J. H., Ma, R. C. W., Song, X. Y., Chan, J. C. N., and So, W. Y. (2014), “Age of Diagnosis, Chronic Hepatitis B Viral Infection and Cardiovascular-Renal Endpoints in Type 2 Diabetes: A 10-year Prospective Cohort Analysis by Structural Equation Modeling,” BMC Public Health, under revision. [1149,1154]

Yin, G., and Ibrahim, J. G. (2005), “A Class of Bayesian Shared Gamma frailty Models With Multivariate Failure Time Data,” Biometrics, 61, 208–216. [1150]

Zeng, D., Lin, D. Y., and Lin, X. (2008), “Semiparametric Transformation Models With Random Effects for Clustered Failure Time Data,” Statistica Sinica, 18, 355–377. [1157]

Zeng, D., Yin, G., and Ibrahim, J. G. (2005), “Inference for a Class of Transformed Hazards Models,” Journal of the American Statistical Association, 100, 1000–1008. [1157]