Regression modeling of semi-competing risks data

Limin Peng, Jason P. Fine, and Rick Chappell
Department of Statistics, University of Wisconsin, Madison, WI

pengl@stat.wisc.edu
fine@biostat.wisc.edu
chappell@biostat.wisc.edu
Regression modeling of semi-competing risks data

Limin Peng, Jason P. Fine, and Rick Chappell

Abstract

Semi-competing risks data are encountered in studies with intermediate endpoints subject to dependent censoring. There has recently been increased attention to this data as distinct from classical competing risks data, in particular, inferences without covariates. In this paper, we incorporate covariates. Instead of modelling hazard functions, we formulate the covariate effects on the marginal and joint survival functions of the events via functional regression models. This includes a novel time-dependent copula model which generalizes parametric copula models. New nonparametric estimators are constructed from nonlinear estimating equations, and are shown to be uniformly consistent and to converge weakly. Inferences for the time-varying covariate effects and copula parameters are developed accordingly. Simulations and an AIDS data analysis illustrate the methodology's practical utility.

KEY WORDS: Empirical processes; Informative dropout; Nonlinear estimating function; Parametric submodel; Time-indexed dependence structure; Varying coefficients.
1. Introduction

Semi-competing risks data (Fine, Jiang, and Chappell, 2001) are encountered when there is a terminating event which potentially censors a nonterminating event which does not prevent subsequent observation of the terminating event. For example, in many clinical trials, the primary analysis may involve an intermediate disease endpoint, but patients may dropout or the study may end prior to the endpoint of interest. If the dropout is informative, then conventional analyses which assume independent censoring may be invalid for the marginal distribution of the disease event, which corresponds to a setting without dropout.

If one only uses information on the time and type of the first event, then this reduced data may be analysed as classical competing risks. There has been much work without covariates. The marginal distributions and dependence structure of the event times are not identifiable nonparametrically (Tsiatis, 1975). Inferences involve either restricting the joint distribution or a sensitivity analysis (Peterson, 1976; Slud, Byar, and Schatzkin, 1988; Klein and Moeschberger, 1988; Link, 1989; Emoto and Matthews, 1990; Zheng and Klein, 1995). It is recognized that the extra information in the semi-competing risks data is helpful in nonparametric estimation (Andersen et al., 1993). However, the marginal distribution of the nonterminating event cannot be identified without further assumptions, similarly to competing risks. Fine et al. (2001) estimated a semiparametric model where the dependence structure satisfies the gamma frailty copula (Clayton, 1978; Oakes, 1989), but the marginal distributions are unspecified. This was extended to other parametric copulas by Wang (2003).

The focus of this paper is regression modeling. A naive proportional hazards analysis which ignores dependent censoring implicitly fits a model for the cause-specific hazard function induced by dropout and not the marginal distribution. The identifiability of bivariate proportional hazards and accelerated lifetime models has been established for competing risks (Heckman and Honore, 1989). These theoretical results have not been translated into practical estimation procedures, owing to complexities in simultaneously estimating all model parameters via maximum likelihood. Recently, the special structure of semi-competing risks data was exploited by Lin, Robins, and Wei (1996). They developed rank-based estimating equations for covariate effects in a bivariate accelerated lifetime model, separately from the dependence structure. Similar rank procedures have not been developed for proportional hazards models, in part, because the hazard model necessitates a maximum likelihood analysis.

Instead of modeling hazard functions, we formulate a class of varying coefficient regression
models for the survival function of the nonterminating event (Fine, Yan, and Kosorok, 2004). This includes extensions of the popular proportional hazards model and the proportional odds models. Mimicking the model for the marginal survival function, we propose to model the dependence structure of the joint survival function of the nonterminating and terminating events with a novel time-dependent copula. This model is more flexible than parametric copulas (Fine et al., 2001; Wang, 2003) and is helpful in assessing the fit of the parametric copulas. The marginal and joint survival models are discussed in Section 2.

The regression modeling strategy facilitates estimation and inference. Computationally simple nonparametric estimators for time-varying covariate effects and time-varying copula parameters are introduced in Section 3. The estimators’ uniform consistency and weak convergence, as well as consistent variance estimators, are described in the text, with proofs in the appendix. Estimation and goodness-of-fit testing for parametric submodels for the parameters in the functional models are presented in Section 4 and permit a formal assessment of time-independent survival models. Simulations and an AIDS data analysis are reported in Section 5. Remarks conclude in Section 6.

2. Data and model

Let $T_1$ be the time to the nonterminating event, $T_2$ be the time to the terminating event, which may dependently censor $T_1$, and $Z^0$ be a $p \times 1$ covariate vector. Let $C$ be a censoring time which is independent of $(T_1, T_2)$ conditionally on $Z^0$. Define $X = T_1 \land T_2 \land C$, $\delta_1 = I(T_1 \leq T_2 \land C)$, $Y = T_2 \land C$ and $\delta_2 = I(T_1 \leq T_2 \land C)$, where the $\land$ is the minimum operator and $I(.)$ is the indicator function. The observed data consist of $n$ replicates of $(X, Y, \delta_1, \delta_2, Z^0)$, denoted by $\{X_i, Y_i, \delta_{i1}, \delta_{i2}, Z_{i0}\}_{i=1}^n$.

In the semi-competing risks setting, $\{Y_i, Z_{i0}, \delta_{i2}\}_{i=1}^n$ can be viewed as standard independently right censored time-to-event data for $T_2$. Hazard regression models for $\text{pr}(T_2 > t|Z^0)$ have been thoroughly studied in this scenario, for example, multiplicative hazards, additive hazards, and accelerated failure-time models. To begin, we do not specify a particular regression model for $T_2$ but do assume that there already exists a method of estimation for the chosen model. The challenge is tailoring a joint model to $(T_1, T_2)$ which permits estimation of covariate effects on $T_1$ in the presence of censoring by $T_2$.

A partial likelihood analysis for $T_1$, which assumes $T_1$ and $T_2$ are independent, could be based on the proportional hazards model, in which the hazard for $T_1$ given $Z^0$ satisfies

$$
\lambda(t|Z^0) = \lambda_0(t) \exp(\beta_0^T Z^0).
$$

(2.1)
The parameter \( \lambda_0(t) \) is an unspecified baseline hazard function and \( \beta_0 \) is a \( p \times 1 \) coefficient vector. The model (2.1) implies that

\[
\Pr(T_1 > t | Z_0) = S(t | Z_0) = \exp[- \exp\{\log \Lambda_0(t) + \beta_0^T Z_0\}]
\]

(2.2)

where \( \Lambda_0(t) = \int_0^t \lambda_0(s) ds \). In practice, restricting the hazard functions associated with two sets of covariates to be proportional over time may be unrealistic. Extending the intensity model (2.1) to time-varying effects has been well studied (Zucker and Karr, 1990; Murphy and Sen, 1991; Fahrmeir and Klinger, 1998; Martinussen, Scheike, and Skovgaard, 2002).

Following Fine et al. (2004), we instead employ the time-varying effects in the survival model (2.2), that is,

\[
S(t | Z) = g\{\theta_0(t)^T Z\},
\]

(2.3)

where \( g(\cdot) \) is a known monotone function, \( Z = \left( \frac{1}{x_0^T} \right) \) and \( \theta_0(t) \) is a \((p+1) \times 1\) vector of unknown time-dependent coefficients. The implied hazard function for \( T_1 \) is \([\dot{g}\{\theta_0(t)^T Z\} + g\{\theta_0(t)^T Z\}] / [\{\partial \theta_0(t)/\partial t\}^T Z]\), where \( \dot{g}(u) = dg(u)/du \).

The model (2.3) defines a rich family of varying coefficient regression models in which \( \theta_0(t) \) is interpretable via the generalized linear model \( g^{-1}\{S(t)\} = \theta_0(t)^T Z \). While the interpretation of time-varying effects in the proportional intensity model (2.1) is straightforward, their interpretation in the survival model is different and relies on the binary regression structure (2.3). For the proportional odds model with \( g = \exp/(1 + \exp) \), the components of \( \theta_0(t) \) are log odds ratios of surviving beyond \( t \) per unit change in the corresponding covariates, just as in logistic regression. If \( g = \exp(-\exp) \), then model (2.3) is the proportional hazards model (2.2) when \( \theta_0(t) = \{\log \Lambda_0(t), \beta_0^T\}^T \). With time-varying effects, the parameter interpretation may be less natural than it is in the proportional odds model.

Fine et al. (2004) developed estimation methods for \( \theta_0 \) with independent right censoring. With semi-competing risks, estimation of \( \theta_0 \) requires a model for the dependence structure of \((T_1, T_2)\), since \( T_2 \) may dependently censor \( T_1 \). Extending the modelling strategy in Fine et al. (2004), we link the joint distribution of \((T_1, T_2)\) to its marginals through a known time-independent copula function \( C(u, v, w) \), where for fixed \( w \), \( C \) satisfies the definition of a copula, and an unknown time-varying parameter \( \alpha_0(s, t) \). It is assumed that

\[
\Pr(T_1 > s, T_2 > t | Z) = C\{\Pr(T_1 > s | Z), \Pr(T_2 > t | Z), \alpha_0(s, t)\}, \text{ for } 0 \leq s \leq t.
\]

(2.4)

For given \( s \) and \( t \), \( C\{\cdot, \alpha_0(s, t)\} \) defines a copula for the binary random variables \( I(T_1 > s) \) and \( I(T_2 > t) \). The model (2.4) generalizes the class of parametric copula models (Clayton and Cuzick, 1985; Hougaard, 1987; Oakes, 1989; Genest and MacKay, 1986) where \( \alpha_0 \) is time-invariant. For example, when \( C(u, v, w) = |u^{1-w} + v^{1-w} - 1|^{1/(1-w)} \) and \( \alpha_0(s, t) = \alpha^*, 0 \leq
$s \leq t$, (2.4) is the gamma-frailty copula on the upper wedge (Day, Bryant, and Lefkopoulou, 1997; Fine et al., 2001). The representation encompasses the dependence structures in Wang (2003).

The time-indexed parameter $\alpha_0(s,t)$ accommodates realistic scenarios where the dependence between $T_1$ and $T_2$ may change over time. In the data analysis in Section 5.2, dropouts at the beginning of follow-up are more likely to be outcome-related, and thus more informative, than dropouts later in the follow-up period. The copula methodology proposed in this paper enables tests for the time-invariance of the association between $T_1$ and $T_2$.

While in theory, the regression model for $\text{pr}(T_2 > t | Z)$ may be specified arbitrarily, it is natural to simplify the developments by assuming that the form of the model for $\text{pr}(T_2 > t | Z)$ is the same as that for $\text{pr}(T_1 > t | Z)$ in (2.3). That is,

$$\text{pr}(T_2 > t | Z) = h \{ \eta_0(t)^T Z \}, \quad (2.5)$$

where $h$ is a known link function and $\eta_0(t)$ is estimable with existing methods. Note that $h$ and $\eta_0(t)$ may not equal $g$ and $\theta_0(t)$. The estimator of $\eta_0(t)$ is denoted by $\hat{\eta}_0(t)$.

3. Nonparametric inference for $(\alpha_0, \theta_0)$

3.1 Nonlinear estimating equations

The methods of Fine et al. (2004) are valid for estimating $\theta_0$ under independent censoring, when the censoring time is always observed. With semi-competing risks, simultaneous estimation of the covariate effects on $T_1$ and the dependence parameters is needed. We now show that the survival models in Section 2 facilitate inferential procedures which are more tractable, theoretically and computationally, than those from hazard models, where likelihood based analyses may be needed.

To estimate $\{ \alpha(t), \theta(t) \}$, consider $A_i \{ \alpha(t), \theta(t), \eta(t), t \} =$

$$V_i \{ \alpha(t), \theta(t), t \} D_i \{ \alpha(t), \theta(t), \eta(t) \} [I(X_i > t) - I(Y_i > t)] \Psi \{ \alpha(t), \theta(t)^T Z_i, \eta(t)^T Z_i \}],$$

where $V_i$ is a scalar weight function, $\alpha(t) = \alpha(t, t), \Psi(u, v, w) = C \{ g(v), h(w), u \}/h(w)$, and $D_i \{ \alpha(t), \theta(t), \eta(t) \} = \partial \Psi \{ \alpha(t), \theta(t)^T Z_i, \eta(t)^T Z_i \}/\partial(\alpha(t), \theta(t), \eta(t))$. Under the assumed models (2.3-5), $A_i \{ \alpha_0(t), \theta_0(t), \eta_0(t), t \}$ has mean zero conditionally on $Z_i$ and $I(Y_i > t)$, since $E \{ I(X_i > t) | Y_i > t, Z_i \} = \Psi \{ \alpha_0(t), \theta_0(t)^T Z_i, \eta_0(t)^T Z_i \}$. The quantity $A_i$ automatically accounts for independent censoring by $C$ by conditioning on $I(Y_i > t)$. Substituting $\hat{\eta}$ for $\eta_0$ and averaging over the $n$ observations yields the estimating equation $U \{ \alpha(t), \theta(t), \hat{\eta}(t), t \} = 0$, where

$$U \{ \alpha(t), \theta(t), \eta(t), t \} = \sum_{i=1}^n A_i \{ \alpha(t), \theta(t), \eta(t), t \}. \quad (3.1.1)$$
The estimator \( \{\hat{\alpha}(t), \hat{\theta}(t)\} \) is the root of \( U \). If the value of \( \alpha_0(t) \) giving the independence of \( T_1 \) and \( T_2 \) is substituted into \( U \), one can show that \( \eta \) vanishes from \( \Psi \) and \( \hat{\theta} \) based on the last \( p \) components of \( U \) equals that in Fine et al. (2004), which assumes independent censoring.

The nonlinear estimating function (3.1.1) separately estimates \( \alpha \) and \( \theta \) at each \( t \), adopting the 'working independence' assumption (Liang and Zeger 1986). The estimators \( \hat{\alpha}(t) \) and \( \hat{\theta}(t) \) are step functions which jump only at observed failure and censoring times. This means that the estimating equation only needs to be solved at finitely many time points. The estimators are calculated in intervals between the jump points by carrying forward their values at the lower endpoint of the interval. When \( V_{i} = 1 \), solving the equation at a given \( t \) is equivalent to minimizing the nonlinear least squares criterion

\[
\sum_{i=1}^{n} [I(X_i > t) - I(Y_i > t)] \Psi \{\alpha(t), \theta(t)^T Z_i, \hat{\eta}(t)^T Z_i\}^2.
\]

Standard statistical software may be used for this optimization, e.g., nlin in SPLUS. This greatly simplifies the computations, which are rather involved for estimating equations which combine information across time, especially since \( \Psi \) is highly nonlinear, and for likelihood analyses, in which the parameters are estimated simultaneously at all \( t \).

3.2 Large sample properties

The theory in Fine et al. (2004) for estimation of \( \theta_0 \) with independent censoring involves estimating equations which are linear in the unknown parameters. It is cannot be extended to the current setting where \( \alpha_0 \) and \( \theta_0 \) are estimated jointly using nonlinear equations and require an estimator for \( \eta_0 \). With known \( \eta_0 \) in (3.1.1), one may establish pointwise consistency and asymptotic normality of \( \{\hat{\alpha}(t), \hat{\theta}(t)\} \) using nonlinear estimating equation techniques. In practice, \( \eta_0(t) \) is replaced by \( \hat{\eta}(t) \), which complicates these pointwise proofs. Stronger uniform convergence results are needed in later sections, leading to further complications.

Assuming uniform consistency of \( \hat{\eta}(t) \), Theorem 1 in the Appendix shows that as \( n \to \infty \) there exists a unique solution to \( U \{\alpha(t), \theta(t), \hat{\eta}(t), t\} = 0 \) in a neighbourhood of \( (\alpha_0, \theta_0) \) which converges to \( (\alpha_0(t), \theta_0(t)) \) in probability, uniformly in \( t \in [l, u] \). It is further shown in Theorem 2 in the Appendix that \( \mathcal{W}(t) = n^{1/2} \{[\hat{\alpha}(t)^T, \hat{\theta}(t)^T]^T - \{\alpha_0(t)^T, \theta_0(t)^T\}^T\} \) converges weakly to a tight Gaussian process under other mild conditions on \( \hat{\eta} \). Suppose \( \sqrt{n} \{\hat{\eta}(t) - \hat{\eta}_0(t)\} \) converges weakly to a tight Gaussian process and may be represented as a sum of iid random functions \( \phi_i(t), i = 1, \ldots, n \). That is, \( \sqrt{n} \{\hat{\eta}(t) - \hat{\eta}_0(t)\} = n^{-1/2} \sum_{i=1}^{n} \phi_i(t) + o_n(t) \), where \( \sup_{t \in [l, u]} o_n(t) \overset{L}{\to} 0 \). Define \( J(t) \) and \( H(t) \) as the asymptotic limits of

\[
\hat{J}(t) = n^{-1} \sum_{i=1}^{n} V_i \{\hat{\alpha}(t), \hat{\theta}(t), t\} [D_i \{\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t)\}] \otimes 2I(Y_i > t), \text{ and}
\]
\[ \hat{H}(t) = n^{-1} \sum_{i=1}^{n} V_i \{ \hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t) \} D_i \{ \hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t) \} D_i^T \{ \hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t) \}^T I(Y_i > t), \]

where \( D_i \{ \alpha(t), \theta(t), \eta(t) \} = \partial \Psi \{ \alpha(t), \theta(t)^T Z_i, \eta(t)^T Z_i \} / \partial \eta(t) \) and for a vector \( u, v = u^T v \).

Then, as \( n \to \infty \), \( W(t) \) follows a Gaussian process with covariance function \( \Sigma(s, t) = E(W_1(s) W_1(t)^T) \), where \( W_1(t) = J(t)^{-1}[A_i \{ \alpha_0(t), \theta_0(t), \eta_0(t), \hat{\phi}_i(t) \} - H(t) \hat{\phi}_i(t)] \).

Let \( \hat{\phi}_i(t) \) be \( \phi_i \) with all unknown quantities replaced by empirical counterparts, such that \( \sup_{t \in [0, \infty]} | \hat{\phi}_i(t) - \phi_i(t) | \to 0 \), for \( i = 1, \ldots, n \). A uniformly consistent estimator of \( \Sigma(s, t) \) is given by \( \hat{\Sigma}(s, t) = n^{-1} \sum_i W_i(s) W_i(t)^T \), where \( W_i(t) = J(t)^{-1}[A_i \{ \hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t), \hat{\phi}_i(t) \} - \hat{H}(t) \hat{\phi}_i(t)] \). Inferences about \( \alpha_0 \) and \( \theta_0 \) at finite time points may be based on a multivariate normal approximation and the estimated covariance function.

In the proof of Theorem 1, model identifiability is assumed in regularity condition (C3), which guarantees that as \( n \to \infty \) \( U \) has a unique root in a neighborhood of \( \{ \alpha_0(t), \theta_0(t) \} \). At a given \( t \), uniqueness occurs if the derivative matrix for \( U \) at \( t \) has a nonsingular limit. It is easy to check that with only one dichotomous covariate, \( J(t) \) is singular for all \( t > 0 \). The problem occurs because, at each \( t \), there are two conditional probabilities identifiable from the data: \( \text{pr}(T_1 > t | T_2 > t, Z = 0) \) and \( \text{pr}(T_1 > t | T_2 > t, Z = 1) \), and three unknown parameters: \( \theta_0(t), \theta_2(t), \alpha(t) \). With either multiple discrete covariates or \( \geq 1 \) continuous covariates having nonzero \( \theta_0 \), there appears to be sufficient variability in the conditional probabilities for nonsingular \( J(t) \). However, formally checking (C3) is difficult analytically and its verification may only be possible on a case-by-case basis.

### 3.3 A special case of \( \hat{\eta} \)

In the Appendix, we give conditions on \( \hat{\eta}(t) \) for the validity of \( \hat{\alpha}(t) \) and \( \hat{\theta}(t) \). Here we sketch the theoretical justification for these properties for the proportional hazards model, which is widely used in practice and is employed in Section 5. The quantities defined below are used in variance estimation for \( \{ \hat{\alpha}, \hat{\theta} \} \) in the simulations and data analysis.

The model for \( T_2 \) is \( P(T_2 > t | Z^0) = \exp\{-\Lambda_0(t) \exp(b_0^T Z^0)\} \), where \( b_0 \) is the \( p \times 1 \) regression coefficient, and \( \Lambda_0(t) \) is the baseline cumulative hazard function, which gives \( \eta_0(t) = (\Lambda_0(t))^{b_0} \). Andersen and Gill (1982) developed asymptotic theory for estimation of \( \{ \Lambda_0(t), \theta \} \) using partial likelihood. Let \( \{ \hat{\Lambda}_0(t), \hat{\theta} \} \) denote the estimator. An estimator for \( \eta_0(t) \) is \( \hat{\eta}(t) = \left( \frac{\log \hat{\Lambda}_0(t)}{\hat{\theta}} \right) \). Uniform consistency and weak convergence of \( \hat{\eta} \) are required in our proofs and may be derived from earlier results (Andersen and Gill, 1982).

Uniform consistency follows directly from the consistency of \( \hat{\theta} \) and uniform consistency of \( \hat{\Lambda}_0(t) \). To demonstrate weak convergence, some notation is required. Let \( N_i(t) = \delta_i(t) I(Y_i \leq t) \) and \( M_i(t) = N_i(t) - \Lambda_0(t) \exp(b_0^T Z_i^0) \). Define \( S^{\{i\}}(b, t) = \frac{1}{n} \sum_{i=1}^{n} (Z_i^0)^{\theta} I(Y_i \geq t) e^{b^T Z_i^0}, j = \)
0, 1, 2. It is standard to assume \( S^{(j)}(b, t) \) converges uniformly to \( s^{(j)}(b) \), \( j = 1, 2, 3 \), for \( t \in [0, T^*] \). Let \( c = s^{(1)} / s^{(0)} \), \( v = s^{(2)} / s^{(0)} - e \otimes 2 \),
\[
G = \int_0^{T^*} v(b_0, t)s^{(0)}(b_0, t)\lambda_{b_0}(t)dt, \quad R(t) = -\int_0^t c(b_0, u)\lambda_{b_0}(u)du.
\]
The functional delta method gives \( \sqrt{n}\{ \hat{\theta}(t) - \theta_0(t) \} \)
\[
= \left( \frac{\sum_{i=1}^n \frac{1}{\lambda_0(t)} \{ R(t)^T \int_0^{t^*} G^{-1} \left\{ Z_i^0 - \frac{s^{(1)}(b_0, s)}{s^{(0)}(b_0, s)} \right\} dM_i(s) + \int_0^t \frac{1}{\lambda_0(s)} dM_i(s) \} \right) + \tau_n^*(t)
\]
\[
= n^{-\frac{1}{2}} \sum_{i=1}^n p_i(t) + \tau_n^*(t)
\]
with \( \sup_{0 \leq t < T^*} \tau_n^*(t) \xrightarrow{p} 0 \). Mimicking Andersen and Gill (1982), \( \{ p_i(t), t \in [l, u] \} \) is Glivenko-
Cantelli and Donsker, and satisfies the conditions in Theorem 2 in the Appendix.

To establish consistency of the variance estimator for \((\hat{\alpha}, \hat{\theta})\), let
\[
\hat{G} = \frac{1}{n} \sum_{i=1}^n \int_0^t \left[ \frac{\sum_{j=1}^n Y_j(s) Z_j^2 e^{btZ_j}}{\sum_{j=1}^n Y_j(s) e^{btZ_j}} - \left\{ \frac{\sum_{j=1}^n Y_j(s) e^{btZ_j}}{\sum_{j=1}^n Y_j(s) e^{btZ_j}} \right\} \right] dN_i(s),
\]
\[
\hat{\tilde{R}}(t) = -\sum_{i=1}^n \int_0^t \frac{\sum_{j=1}^n Y_j(s) Z_j e^{btZ_i} e^{btZ_j}}{\{ \sum_{j=1}^n Y_j(s) e^{btZ_j} \}^2} dN_i(s), \quad \hat{\tilde{p}}^{(1)}_i = \hat{G}^{-1} \left[ \delta_{2i} \left\{ Z_{i}^0 - \frac{S^{(1)}(\hat{b}, Y_i)}{S^{(0)}(\hat{b}, Y_i)} \right\} \right],
\]
\[
\hat{\tilde{p}}^{(2)}_i = \frac{1}{\lambda_0(t)} \left[ \hat{\tilde{R}}(t)^T \hat{G}^{-1} \left\{ \delta_{2i} \left\{ Z_{i}^0 - \frac{S^{(1)}(\hat{b}, Y_i)}{S^{(0)}(\hat{b}, Y_i)} \right\} \right\} + \left\{ \delta_{2i} I(Y_i \leq t) \frac{1}{S^{(0)}(\hat{b}, Y_i)} - \hat{\omega}_b(t) \right\} \right].
\]
The consistency of \( \hat{G} \) and the uniform consistency of \((\hat{\tilde{R}}, \hat{\omega}_b)\) imply \( \sup_{t \in [0, T^*]} \| \{ \hat{\tilde{p}}^{(2)}(t), \hat{\tilde{p}}^{(1)}(t) \} - \{ p^{(2)}(t), p^{(1)}(t) \} \| \xrightarrow{p} 0 \), and the uniform consistency of \( \hat{\Sigma} \) follows.

3.4 Hypothesis testing

We use \((\hat{\alpha}, \hat{\theta})\) from Section 3.2 to construct nonparametric test statistics for the null hypothesis \( H_0 : C(t)\theta(t) = c(t) \), where \( C(t) \) is a \( r \times (p + 2) \) matrix and \( c(t) \) is a \( r \times 1 \) vector. The issue is how to combine the information across time.

Motivated by the Wald test, one can choose \( K \) time points, \( h_1, \ldots, h_K \) and test \( H_0 \) at these time points. Let \( \hat{\beta}^* = \left\{ \left( \hat{\alpha}(h_1) \right)^T, \ldots, \left( \hat{\alpha}(h_K) \right)^T \right\}^T \), \( C^* = \text{diag}\{ C(h_1), \ldots, C(h_K) \} \), and \( c^* = \{ c(h_1)^T, \ldots, c(h_K)^T \}^T \). The test statistic is given by \( T_1 = (C^* \hat{\beta}^* - c^*)^T (C^* \hat{\Sigma}^* C^* - T) (C^* \hat{\beta}^* - c^*) \), where \( \hat{\Sigma}^* \) is the estimated variance matrix of \( \hat{\beta}^* \), which may be derived from Theorem 2. Under conditions ensuring pointwise convergence and asymptotic normality of \( \left( \frac{\hat{\alpha}(t)}{\theta(t)} \right) \), the limiting distribution of \( T_1 \) under \( H_0 \) is \( \chi^2_{rK} \).
Another strategy is to integrate the contrast across time. This gives the test statistic

$$T_2 = \left[ \int_t^u \{ C(t) \hat{\alpha}(t) \} - c(t) \} \Xi(t) dt \right]^{T} \hat{\Sigma}_2^{-1} \left[ \int_t^u \{ C(t) \hat{\alpha}(t) \} - c(t) \} \Xi(t) dt \right]$$

where $\Xi$ is a weight function, and

$$\hat{\Sigma}_2 = n^{-2} \sum_{i=1}^{n} \left( \int_t^u C(s) \hat{W}_i(s) \Xi(s) ds \right) \otimes 2.$$

Under the conditions ensuring uniform convergence of the estimators in Theorem 2, the limiting distribution of $T_2$ under $H_0$ is $\chi^2$.

A third possibility is a supremum-norm test based on

$$T_3 = \sup_{t \in [t, u]} \left\| \{ C(t) \hat{\alpha}(t), \hat{\theta}(t) \} - c(t) \} \xi(t) \{ C(t)^T \hat{\Sigma}(t, t) C(t) \}^{-1} \left\{ C(t) \hat{\alpha}(t), \hat{\theta}(t) \} - c(t) \right\|.$$

The limiting distribution of $T_3$ is much more complicated than that of $T_1$ or $T_2$ and resampling is needed to approximate the distribution of $T_3$, making this test computationally intensive.

Regarding power, for $T_1$, the choice of time points is subtle since it may miss deviations from the null and different sets of time points may give different conclusions. For $T_2$, care is needed when $C(t) \{ \hat{\alpha}(t), \hat{\theta}(t) \} - c(t)$ changes sign in the time interval $[t, u]$. Deviations from $H_0$ may cancel in the integral. To increase power, one may consider data driven weights which accentuate differences from the null. The statistic $T_3$ to departures from $H_0$, but is known to have reduced power because of a lack of specificity to particular alternatives.

4. Parametric modeling of $(\alpha_0, \theta_0)$

4.1 Estimation

In practice, it may be of interest to explore the parametric forms of covariate effects and association parameters. We consider the model $L^T(t) \{ \alpha_0(t)^T, \theta_0(t)^T \}^T = q(\zeta_0, t)$, where $L(t)$ is a $r \times (p + 2)$ vector, $q$ is a known function, and $\zeta_0$ is a finite dimensional parameter. This general framework includes situations where either a certain covariate effect or copula parameter follows a polynomial curve, e.g., time-independent. It has been used in other functional data analyses where parametric models for varying coefficients are of interest.

Under the assumed model, an estimator of $\zeta_0$, say $\hat{\zeta}$, may be defined as the minimiser of the least square criterion $\int_t^u \{ L^T(t) \{ \hat{\alpha}(t), \hat{\theta}(t) \} - q(\zeta, t) \}^2 \bar{\Xi}(t) dt$, where $\bar{\Xi}$ is a nonnegative weight function. The estimator $\hat{\zeta}$ is a solution of $\hat{U}(\hat{\alpha}, \hat{\theta}, \zeta) = 0$, where

$$\hat{U}(\alpha, \theta, \zeta) = \int_t^u \hat{q}(\zeta, t) \{ L^T(t) \{ \alpha(t), \theta(t) \} - q(\zeta, t) \} \bar{\Xi}(t) dt,$$
and \( \dot{q}(\zeta, t) = \partial q(\zeta, t) / \partial \zeta \). We can show that under mild assumptions on the parameter space for \( \zeta_0 \), the function \( q \), and \( \tilde{\zeta} \), \( \zeta \) is consistent for \( \zeta_0 \). Furthermore, a Taylor expansion of \( \tilde{U}(\tilde{\alpha}, \tilde{\beta}, \zeta) \) in \( \zeta \) about \( \zeta_0 \) yields that

\[
n^{1/2}(\dot{\zeta} - \zeta_0) = I(\zeta_0)^{-1} \int_{t_i}^{\tau} \dot{q}(\zeta_0, t) L(t)^T \sqrt{n} \left( \frac{\dot{\alpha}(t) - \alpha_0(t)}{\dot{\theta}(t) - \theta_0(t)} \right) \tilde{\varepsilon}(t) dt + o_p(1),
\]

where \( I(\zeta) = \int_{t_i}^{\tau} \dot{q}(\zeta, t) \dot{q}(\zeta, t)^T \tilde{\varepsilon}(t) dt \). Let

\[
u_t(\alpha, \theta, \eta, \zeta) = I(\zeta)^{-1} \int_{t_i}^{\tau} \dot{q}(\zeta, t) L(t)^T J(t)^{-1} [A_t \{ \alpha(t), \theta(t), \eta(t), t \} - H(t) \phi_t(t)] \tilde{\varepsilon}(t) dt
\]

and \( \bar{\nu}_i \) with \( J \) and \( H \) replaced by \( \bar{J} \) and \( \bar{H} \), respectively. By the asymptotic equivalence of \( \sqrt{n} \left( \frac{\hat{\alpha}(t) - \alpha_0(t)}{\hat{\theta}(t) - \theta_0(t)} \right) \) and \( n^{-1/2} \sum_{i=1}^{n} W_i(t) \) developed in the proof of Theorem 2 in the Appendix, it follows that \( \sqrt{n} \left( \dot{\zeta} - \zeta_0 \right) \) is asymptotically equivalent to \( n^{-1/2} \sum_{i=1}^{n} \bar{\nu}_i(\alpha_0, \theta_0, \eta_0, \zeta_0) \) and has a limiting normal distribution with mean zero and variance \( \Gamma = E \{ \bar{\nu}_i(\alpha_0, \theta_0, \eta_0, \zeta_0) \} \). A consistent plug-in estimator of \( \Gamma \) is \( \hat{\Gamma} = n^{-1} \sum_{i=1}^{n} \bar{\nu}_i(\hat{\alpha}, \hat{\theta}, \hat{\eta}, \hat{\zeta}) \). Detailed proofs of the asymptotic properties of \( \dot{\zeta} \) are available from the authors upon request.

### 4.2 Goodness-of-fit testing

Goodness-of-fit tests are important for assessing the adequacy of the parametric model. The null hypothesis is that the model is correctly specified; that is, \( H_0^* : \ L(t)^T \left( \frac{\hat{\alpha}(t)}{\hat{\theta}(t)} \right) = q(\zeta_0, t) \). It is straightforward to construct test statistics similarly to those in Section 3.4 using \( S(t) = L(t)^T \{ \hat{\alpha}(t)^T, \hat{\theta}(t)^T \}^T - q(\zeta, t) \). The primary difficulty is accounting for the variability in estimating \( \zeta_0 \). Under \( H_0^* \), expanding in \( \zeta_0 \) around \( \tilde{\zeta} \) gives

\[
S_n(t) = L(t)^T \left( \frac{\hat{\alpha}(t) - \alpha_0(t)}{\hat{\theta}(t) - \theta_0(t)} \right) - \dot{q}(\zeta_0, t)^T (\dot{\zeta} - \zeta_0) + o_p(1).
\]

Substituting \( \sqrt{n} \left( \frac{\hat{\alpha}(t) - \alpha_0(t)}{\hat{\theta}(t) - \theta_0(t)} \right) \) with \( n^{-1/2} \sum_{i=1}^{n} W_i(t) \), \( \sqrt{n} S_n(t) \) is asymptotically equivalent to \( n^{-1/2} \sum_{i=1}^{n} j_i(t), \) where \( j_i(t) = L(t)^T J(t)^{-1} [A_t \{ \alpha_0(t), \theta_0(t), \eta_0(t), t \} - H(t) \phi_t(t)] - \dot{q}(\zeta_0, t)^T \). Using the properties of \( \hat{\alpha}, \hat{\theta}, \hat{\zeta} \), one can establish that \( S_n(t) \) converges weakly to a Gaussian process with covariance function \( \Omega(s, t) \), which is the asymptotic limit of \( \Omega(s, t) = n^{-1} \sum_{i=1}^{n} j_i(s) j_i(t)^T \). Here \( j_i \) is \( j_i \) with \( J, H, \alpha, \theta, \eta, \zeta \) replaced by \( \bar{J}, \bar{H}, \hat{\alpha}, \hat{\theta}, \hat{\eta}, \hat{\zeta} \).

The goodness-of-fit tests are as follows. Let \( h_1 < h_2 < \cdots < h_K \) be \( K \) distinct time points, \( S_n = \{ S_n(h_1), \ldots, S_n(h_K) \} \) and \( \Omega^* \) be the estimated covariance of \( S_n^* \). The limiting distribution of the Wald-type statistic \( T_1^* = \sqrt{n} S_n^* \) is \( \chi^2_K \) under \( H_0^* \). The integral statistic is \( T_2^* = \left\{ \int_{t_i}^{\tau} S_n(t) ^T \hat{\Omega}(t) dt \right\} \) where \( \hat{\Omega}(t) \) is a nonnegative weight function and \( \Omega_2 = n^{-2} \sum_{i=1}^{n} \left\{ \int_{t_i}^{\tau} j_i(t) ^T \hat{\Omega}(t) dt \right\} \). It can be shown that \( T_2^* \) is asymptotically
distributed under $H_0^*$. The supremum-norm statistic is $T_3^* = \sup_t |S_n(t)^2 \hat{\Omega}(t, t)^{-1}|$. The limiting distribution of $T_3^*$ has very complicated analytical form but may be approximated numerically via resampling. The strengths and weaknesses of $T_1^*$, $T_2^*$ and $T_3^*$, are closely related to those of $T_1$, $T_2$ and $T_3$, which are discussed in Section 3.4.

5. Numerical studies

5.1 Simulations

Simulation studies are conducted to evaluate the performance of the estimators and test statistics. We generate $\log(T_{1j}/3) = -\theta_2 Z_0^j + e_{1j}$ and $\log(T_{2j}/3) = -\eta_2 Z_0^j + e_{2j}$ for $j = 1, \ldots, n$, where $\theta_2 = 1, \eta_2 = 0$ or 0.2, $Z_0^j$ are independent normally distributed variates, with mean 1 and variance 0.5, constrained to the interval $[0, 2]$. We take $\Pr(e_{1j} > x) = F(x) = \exp\{-\exp(x)\}$, for $j = 1, 2$, which gives proportional hazards models for $T_{1j}$ and $T_{2j}$. The joint distribution of $(e_{1j}, e_{2j})$ is defined by the gamma frailty copula, that is, $\Pr(e_{1j} > t_1, e_{2j} > t_2) = [F(t_1)^{1-a} + F(t_2)^{1-a} - 1]^{1/(1-a)}$. In the simulations, $\alpha = 1.5$ and the sample size $n$ is either 250 or 500. Independent censoring times $C_j$ are generated from Unif(1, 10). When $\eta_2 = 0$, the censoring percentages for $T_1$ and $T_2$ are 27% and 23% respectively. With $\eta_2 = 0.2$, these percentages are 31% and 18%.

Following Section 3.4, an estimator for $\eta_0(t)$ in the proportional hazards model for $T_2$ may be obtained with standard software packages, e.g., coxph in SPLUS. The coefficient vector for $T_1$ and the copula parameter under our simulation scheme are $\alpha(t) = \alpha, \theta_0(t) = \left\{\theta_{01}(t), \theta_{02}(t)^T\right\}^T = \left\{\log(t/3), \theta_2^T\right\}^T$, for $t > 0$. To examine the efficiency of estimating function (3.1.1), we focus on $\theta_{02}(t)$, the effect of $Z_0$ on the survival probability of $T_2$ at time $t$. The weight function $V_t$ is set to 1 and the root is obtained as the minimizer of (3.1.2).

It is found that variance estimation using $\hat{\Sigma}$ may be unstable with $n = 250$ and 500, which may be due to the nonlinearity of the model. We propose an adjusted variance estimator, $\hat{\Sigma}(s, t)$, in which $\hat{J}(t)$ and $\hat{H}(t)$ are replaced by $\bar{J}(t)$ and $\bar{H}(t)$, where

$$\bar{J}(t) = \frac{1}{n} \sum_{i=1}^n D_i \{\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t)\} \otimes^2 I(Y_i > t)$$

$$= \nabla_{\alpha(t)} D_i \{\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t)\} \left[I(X_i > t) - I(Y_i > t) \Psi\{\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t)\}\right]$$

$$\bar{H}(t) = \frac{1}{n} \sum_{i=1}^n D_i \{\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t)\} D_i^* \{\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t)\}^T I(Y_i > t)$$

$$= \nabla_{\eta} D_i \{\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t)\} \left[I(X_i > t) - I(Y_i > t) \Psi\{\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t)\}\right].$$
The quantities $\hat{J}(t)$ and $\hat{H}(t)$ account for a term in the Taylor expansion in the proof of Theorem 2 which is negligible asymptotically but may not be with moderate sample sizes. The variance estimator using $\hat{J}$ and $\hat{H}$ is thus more stable than that using $\hat{J}$ and $\hat{H}$.

In Table 1, we report biases, empirical variances, model based variances and coverage probabilities of 95% confidence intervals from $\hat{\theta}_2(t)$ at time points dividing $[0.22, 1.22]$ into five equal length pieces. The entries are based on 1000 simulated datasets. The point estimates of $\theta_2$ are virtually unbiased, the model based variances are reasonably close to the empirical variances, and the 95% confidence interval has accurate coverage probability.

Next, we test the hypothesis $H_0: \theta_{20}(t) = 1$ by using $T_1$ and $T_2$ from §2.4. The statistic $T_1$ is constructed at time points $(0.47, 0.97)$ and $(0.47, 0.72, 0.97)$, respectively; $T_{12}$ and $T_{13}$ denote the corresponding test statistics. Setting $[l, u] = [0.22, 1.22]$, $T_2$ is computed at the same time as we estimate the parametric submodel $\theta_{20}(t) = \theta_2$. Apparently, $T_2$ is equal to the square of the difference between $\hat{\theta}_2$ and one, the hypothesized value for $\theta_{20}(t)$, divided by the estimated variance of $\hat{\theta}_2$. In obtaining $\hat{\theta}_2$, we take the weight function in the least square criterion in Section 4.1 to be the inverse of the estimated variance of $\hat{\theta}_2(t)$. Table 2 presents the biases, empirical variances and model based variances for the least squares estimator of $\theta_2$, as well as the rejection rates from $T_{12}$, $T_{13}$ and $T_2$ at significance level 0.05. The tests reject at rate close to the nominal rate, the model based variances and empirical variances are in close agreement, and second stage estimator of $\theta_2$ performs well.

The power of the test statistics for no covariate effect is now investigated. We take the null to be $H_0: \theta_{22}(t) = 0$. The last three columns of Table 2 give the power of $T_{12}$, $T_{13}$ and $T_2$. It appears that all tests have substantial power to reject the null hypothesis that the covariate $Z^0$ has no effect when $\theta_2 = 1$.

5.2 Real example

In AIDS Clinical Trial Group (ACTG) Study 364 (Albrecht et al., 2001), the time to first virologic failure, $T_1$, is a primary endpoint and is subject to dependent censoring by dropout before study completion at time $T_2$. There is also administrative censoring at time $C$ for those subjects observed until the end of the follow-up period.

There were 195 patients randomly assigned to three antiretroviral treatments: nelfinavir (NFV), efavirenz (EFV), and NFV+EFV. Data from each treatment arm were analysed separately by Jiang, Chappell, and Fine (2003) as semi-competing risks, without adjusting for other covariates. In Albrecht et al. (2001), there was interest baseline HIV RNA and whether 3TC is a new nucleoside reverse transcriptase inhibitor (NRTI). These covariates had
significant effects in conventional proportional hazard analyses of $T_1$ which do not account for dependent censoring by $T_2$. We reanalyse the data using the modeling framework in Section 2.

Two binary covariates are defined for treatment. Let $Z_1 = 1$ if a patient was assigned EFV and 0 otherwise, and let $Z_2 = 1$ if a patient was assigned NFV+EFV, and 0 otherwise. The covariate $Z_3$ equals 1 if the patient received 3TC as a new NRTI in ACTG 364 study and 0 otherwise. The continuous covariate $Z_4$ equals $\log_{10}$ baseline RNA. Let $Z^0 = (Z_1, Z_2, Z_3, Z_4)^T$. One subject is missing $Z_3$ and is excluded from the analysis. During the two-year follow-up period, 101 patients were observed to experience virologic failures, while the other 93 patients were censored, 82 administratively by $C$ and 11 due to dropout by $T_2$. While the dropout rate does not appear to be great, there are differences amongst the treatment arms (0 on NFV, 1 on EFV, 10 on NFV+EFV) and ignoring these differences may bias the analysis if dropout is correlated with virologic failure.

We fit a proportional hazards models to $T_2$ with $h = \exp(-\exp)$ in (3.1.1). For $T_1$, we took $g = \exp(-\exp)$ in (2.3) with $\theta_0(t) = \{\theta_{00}(t) \theta_{01}(t) \theta_{02}(t) \theta_{03}(t) \theta_{04}(t)\}^T$, which gives a proportional hazards model when $\theta_{0j}(t), j = 1, \ldots 4$ are time-invariant. This enables a comparison with a partial likelihood analysis which ignores the potentially dependent censoring of $T_1$ by $T_2$. The dependence structure $C(u, v, w) = \{u^{1-\exp(w)} + v^{1-\exp(w)}\}^{-1/(1-\exp(w))}$ in (2.4), which extends the gamma-frailty copula with constant predictive hazard ratio, $\exp(\alpha)$, where $\alpha(t) = \hat{\alpha}$ (Day et al., 1997). For $\hat{\alpha} > 0$ and $< 0$, there is positive and negative correlation between $T_1$ and $T_2$, respectively, with $\hat{\alpha} = 0$ implying independence of $T_1$ and $T_2$.

Under (2.4), independence of $T_1$ and $T_2$ implies $\alpha(t) = 0$, all $t$. To test this hypothesis, $T_1$ is calculated at three, four and five time points, which are chosen to partition $[1, u]$ into four, five and six equal length segments, and $T_2$ is computed with unit weights. In the sequel, $u$ and $l$ are the first and third quartiles of $X_t$, which are 156 and 672, respectively. The statistics $T_{13} = 12.430$, $T_{14} = 16.545$, $T_{15} = 39.879$, and $T_2 = 9.870$, with the $p$ values all $< 0.01$, providing strong evidence for correlation between $T_1$ and $T_2$.

Fig. 1 displays $\hat{\alpha}$ and $\hat{\theta}_j, j = 1, 2, 3, 4$ along with pointwise 95% confidence intervals. Observe that there is a strong positive dependence at early time points which decreases linearly to one year and then levels at $\hat{\alpha}(t) \approx 0.4$ at later $t$. When the true model is not a constant model, goodness-of-fit tests which integrate $\hat{\alpha}(t) - \zeta_0$ over $t \in [z, u]$ may have low power when positive and negative deviations cancel. To address this issue, we also compute $T_2^{*}(t)$ based on two other weight functions, $\tilde{\zeta}(t) = I(t < (u + l)/2)$ and $\check{\zeta}(t) = I(t > (u + l)/2)$. The resultant test statistics are denoted by $T_2^{*}(t)$ and $T_2^{**}(t)$, while
\( T_{21}(t) \) denotes the test with weight \( \tilde{z}(t) = 1 \). Testing the constant model yields \( T_{13} = 2.499 \), \( T_{14} = 5.022 \), \( T_{15} = 4.473 \), \( T_{21} = 4.415 \), \( T_{22} = 3.611 \) and \( T_{23} = 1.825 \), with \( p \) values 0.475, 0.285, 0.483, 0.036, 0.057, 0.177, respectively. Notice that the integrated difference tests capture the observed decreasing trend at early time points, while the finite time points test is less sensitive, in part because information at later times is given the same weight as that at earlier times. Rejecting the constant model suggests that the time-independent gamma frailty copula may not adequately describe the association between \( T_1 \) and \( T_2 \).

To test for covariate effects in the presence of dependent censoring, \( T_1 \) is calculated at three and four time points, along with \( T_2 \), as described above for \( \alpha(t) \). Table 3 gives the values of these tests and their \( p \) values. The results of the integral tests are quite similar to those from naively fitting a proportional hazards models on \( T_1 \) (see Table 4), although \( T_1 \) is somewhat conservative. The combination therapy has reduced rate of virologic failure relative to the single line treatments. Individuals for whom 3TC is not a new nucleoside and those with elevated baseline RNA are at increased risk of failure. All covariate effects are fairly constant in Fig. 1. Constant parametric models \( \theta_{0i}(t) = \theta_{0i} \) are fitted and evaluated with goodness-fit-tests. Table 5 reports parameter estimates and tests from \( T_{1*} \) and \( T_{2*} \), where \( T_{1*} \) is based on three, four and five time points, as above. The large \( p \) values in Table 5 suggest that time-independent coefficients may be adequate. The results from (2.1.3) are similar to those from fitting the Cox model directly via partial likelihood, even though \( T_1 \) and \( T_2 \) are dependent. This is partially explained by the low dependent censoring rate.

6. Remarks

The time-dependent marginal and bivariate survival models are motivated, in part, by their convenience in developing inferential procedures with semi-competing risks data, which have been rather challenging with hazard based modeling approaches. As demonstrated in the AIDS data analysis, the time-dependent copula model may be very useful in diagnosing the lack-of-fit of more restrictive parametric copula models. This modeling framework is generic and should be broadly applicable with other censoring patterns, like bivariate right censoring. Such applications are a topic for future research.

The computationally simple estimation procedure in Section 3.1 may be viewed as indirect, if the parametric submodels are of primary interest. That is, one first estimates \((\hat{\alpha}, \hat{\theta})\) and then explores the submodels using these estimators in the second stage of inferences. However, the procedure is robust, in the sense that parametric inferences for a particular component of \((\alpha, \theta)\) do not depend on correct specification of parametric models for the other
components, for example, time-independent gamma frailty copula for $\alpha$. Direct estimation of a mixed model in which certain components are assumed to satisfy more restrictive models would complicate the analysis and would no longer be robust. Separate estimation could not be carried out at each time point, similarly to standard likelihood analyses.

REFERENCES


**APPENDIX**

*Theorems and proofs*

Define

\[
K_i(\tau, \kappa, \zeta, \alpha, \theta, \eta, t) = V_i(\tau(t), \kappa(t), t) D_i(\tau(t), \kappa(t), \zeta(t)),
\]

\[
[I(X_i > t) - I(Y_i > t)] \Psi \{\alpha(t), \theta(t) \Psi N(t) \Psi Z(t) \Psi Z_i},
\]

\[
L_i(\tau, \kappa, \zeta, \alpha, \theta, \eta, t) = V_i(\tau(t), \kappa(t), t) D_i(\tau(t), \kappa(t), \zeta(t)) D_i(\alpha(t), \theta(t), \eta(t))^T I(Y_i > t),
\]

\[
Q_i(\tau, \kappa, \zeta, \alpha, \theta, \eta, t) = V_i(\tau(t), \kappa(t), t) D_i(\tau(t), \kappa(t), \zeta(t)) D_i(\alpha(t), \theta(t), \eta(t))^T I(Y_i > t),
\]

18
and $L(\tau, \kappa, \zeta, \alpha, \theta, \eta, t) = E\{L_i(\tau, \kappa, \zeta, \alpha, \theta, \eta, t)\}$. The regularity conditions are:

(C1) $\sup_{t \in [l, u]} |\alpha_0(t)| < \infty$, $\sup_{t \in [l, u]} ||\theta_0(t)|| < \infty$, $\sup_{t \in [l, u]} ||\eta_0(t)|| < \infty$, and $Z$ is bounded with probability 1.

(C2) $\Psi(u, v, w)$ and all components of $\partial\Psi(u, v, w)/\partial(u v w)$ are Lipschitz continuous.

(C3) $\inf_{t \in [l, u]} \text{eigmin} J(t) > 0$, where eigmin is the minimum eigenvalue of a matrix.

(C4) Random weight functions $V_i (i = 1, \ldots, n)$ are independent of $X_i (i = 1, \ldots, n)$ and $Y_i (i = 1, \ldots, n)$. For all bounded $A \subset \mathcal{R}$ and $B \subset \mathcal{R}^{p+1}$, the class of random function $\{V_1(a, b, t), a \in A, b \in B, t \in [l, u]\}$ is bounded below and above by positive constants and is a Donsker class (van der Vaart and Wellner, 1996). In particular, if $V_1(a, b, t) = V(a, b, t, Z_t)$ where $V$ is a nonrandom function and is Lipschitz continuous, then function class $\{V_1(a, b, t), a \in A, b \in B, t \in [l, u]\}$ is Donsker.

(C5) All components of $\partial^2\Psi(u, v, w)/\partial(u v w)\partial(u v w)^T$ are Lipschitz continuous. For all bounded $A \subset \mathcal{R}$ and $B \subset \mathcal{R}^{p+1}$, the class of random functions $\{\partial V_1(a, b, t)/\partial(a b t)^T, a \in A, b \in B, t \in [l, u]\}$ is bounded and is a Donsker class.

**Theorem 1.** Suppose the true model has the form in (2.3-5), and conditions (C1-4) hold. If $\hat{\eta}(t)$ is a uniformly consistent estimator of $\eta_0(t)$ for $t \in [l, u]$, then for $n$ large enough, there exists a uniformly bounded solution of $U\{\alpha(t), \theta(t), \hat{\eta}(t), t\} = 0$, $(\hat{\theta}(t))$, such that $\sup_{t \in [l, u]} \left\|\left(\hat{\theta}(t) - \theta_0(t)\right)\right\| \to 0$ in probability.

**Proof of Theorem 1:** Define $\mathcal{G} = \{K_i(\tau, \kappa, \zeta, \alpha, \theta, \eta) : \tau, \alpha \in \ell_c^\infty([l, u]), \kappa, \zeta, \theta, \eta \in \ell_c^\infty([l, u])^{p+1}\}$. The class $\mathcal{G}$ is Glivenko-Cantelli and Donsker since $\{I\{(t, \infty)\} : t \in [l, u]\}$ is a Donsker class; since $\{\theta(t)^TZ_i, \theta \in \ell_c^\infty([l, u])^{p+1}, t \in [l, u]\}$ and $\{\eta(t)^TZ_i, \eta \in \ell_c^\infty([l, u])^{p+1}, t \in [l, u]\}$ are both equivalent to the Donsker class $\{b^TZ_i, b \in [-c, c]^{p+1}\}$ since Donsker property is preserved under Lipschitz transformation, sum and product operations; since every Donsker class is a Glivenko-Cantelli class in probability.

For any $\hat{\alpha} \in \ell_c^\infty([l, u])$, $\hat{\theta} \in \ell_c^\infty([l, u])^{p+1}$, one can easily show that

$$n^{-1}U(\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t), t) = n^{-1} \sum_{i=1}^n \left(K_i(\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t), t) - V_i(\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t), t)\right)$$

$$I(Y_i > t) \left[\Psi(\hat{\alpha}(t), \hat{\theta}(t)^TZ_i, \hat{\eta}(t)^TZ_i) - \Psi(\alpha_0(t), \theta_0(t)^TZ_i, \eta_0(t)^TZ_i)\right]$$

$$= n^{-1} \sum_{i=1}^n \left(K_i(\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t), t) - V_i(\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t), t)\right) D_i(\hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t)) I(Y_i > t)$$
\[
\begin{aligned}
&\left[ D_t\{\tilde{\alpha}(t), \tilde{\theta}(t), \tilde{\eta}(t)\}^T : \begin{pmatrix}
\tilde{\alpha}(t) - \alpha_0(t) \\
\tilde{\theta}(t) - \theta_0(t)
\end{pmatrix} + D_t^*\{\tilde{\alpha}(t), \tilde{\theta}(t), \tilde{\eta}(t)\}^T \{\tilde{\eta}(t) - \eta_0(t)\} \right] \\
&= n^{-1} \sum_{i=1}^{n} \left\{ K_i(\tilde{\alpha}(t), \tilde{\theta}(t), \tilde{\eta}(t), \alpha_0(t), \theta_0(t), \eta_0(t), t) - V_i(\tilde{\alpha}(t), \tilde{\theta}(t), t) D_t\{\tilde{\alpha}(t), \tilde{\theta}(t), \eta_0(t)\} I(Y_i > t) \right\} \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} D_t\{\tilde{\alpha}(t), \tilde{\theta}(t), \eta_0(t)\}^T \begin{pmatrix}
\tilde{\alpha}(t) - \alpha_0(t) \\
\tilde{\theta}(t) - \theta_0(t)
\end{pmatrix} + \varepsilon_n(t)
\end{aligned}
\]

where \(\{\tilde{\alpha}(t)^T \tilde{\theta}(t)^T \tilde{\eta}(t)^T\}^T\) is on the line segment between \(\{\tilde{\alpha}(t)^T \tilde{\theta}(t)^T \tilde{\eta}(t)^T\}^T\) and \(\{\alpha_0(t)^T \theta_0(t)^T \eta_0(t)^T\}^T\). Since \(\tilde{\eta}(t)\) is a uniformly consistent estimator, \(\sup_{t \in [l, u]} \|\varepsilon_n(t)\| \xrightarrow{P} 0\) follows from uniform boundedness of \(V_i\) by (C4), uniform boundedness of \(\alpha_0(t), \theta_0(t), \eta_0(t)\) by (C1), and the Lipschitz continuity of \(D_t\) and \(D_t^*\) implied by (C2).

Since \(G\) is Glivenko-Cantelli and \(E\{K_i(\tilde{\alpha}, \tilde{\theta}, \tilde{\eta}, \alpha_0, \theta_0, \eta_0, t)\} = 0\) for all \(\tilde{\alpha} \in \{L^\infty_{\xi}(\{l, u\})\}\) and \(\tilde{\theta}, \tilde{\eta} \in \{L^\infty_{\xi}(\{l, u\})\}^{p+1}\), it follows that \(\sup_{t \in [l, u]} \|n^{-1} \sum_{i=1}^{n} K_i(\tilde{\alpha}, \tilde{\theta}, \tilde{\eta}, \alpha_0, \theta_0, \eta_0, t)\| \xrightarrow{P} 0\).

We can similarly establish the Glivenko-Cantelli and Donsker properties for \(G^* = \{L_1(\tau, \kappa, \xi, \alpha, \theta, \eta) : \tau, \alpha, \kappa, \xi, \theta, \eta \in \{L^\infty_{\xi}(\{l, u\})\}^{p+1}\}\) and \(G^{**} = \{Q_i(\tau, \kappa, \xi, \alpha, \theta, \eta) : \tau, \alpha, \kappa, \xi, \theta, \eta \in \{L^\infty_{\xi}(\{l, u\})\}^{p+1}\}\). Then for \(\tau, \alpha, \kappa, \xi, \theta, \eta \in \{L^\infty_{\xi}(\{l, u\})\}^{p+1}, \|n^{-1} \sum_{i=1}^{n} V_i(\tau(t), \kappa(t), \xi(t), \theta(t), \eta(t))\| \xrightarrow{P} 0\) uniformly in \(t \in [l, u]\). Therefore,

\[
n^{-1} U\{\tilde{\alpha}(t), \tilde{\theta}(t), \tilde{\eta}(t), t\} = -L(\tilde{\alpha}, \tilde{\theta}, \tilde{\eta}, \alpha_0, \tilde{\theta}, \eta_0) \begin{pmatrix}
\tilde{\alpha}(t) - \alpha_0(t) \\
\tilde{\theta}(t) - \theta_0(t)
\end{pmatrix} + \varepsilon_n^*(t)
\]

with \(\sup_{t \in [l, u]} \|\varepsilon_n^*(t)\| \xrightarrow{P} 0\).

By condition (C3), boundedness of \(V_i\) and continuity of \(D_t, D_t^T\), there exists a small neighbourhood of \((\alpha_0, \theta_0, \eta_0, \alpha_0, \theta_0, \eta_0)\) in \([L^\infty_{\xi}(\{l, u\}) \times \{L^\infty_{\xi}(\{l, u\})\}^{p+1} \times \{L^\infty_{\xi}(\{l, u\})\}^{p+1}]^2\), inside of which, \(\inf_{t \in [l, u]} \text{eigmin} L(\tau, \kappa, \xi, \alpha, \theta, \eta, t)\) is bounded below by a positive constant \(k\). For \(n\) large enough so that \(\varepsilon_n^*(t)\) small enough, existence of a unique, uniformly bounded solution to \(U\{\alpha(t), \theta(t), \eta(t), t\} = 0\) in this neighbourhood is implied by the inverse function theorem (Goffman, 1965). The uniform consistency follows from

\[
0 = \|n^{-1} U\{\tilde{\alpha}(t), \tilde{\theta}(t), \tilde{\eta}(t), t\}\| \geq k \left\| \begin{pmatrix}
\tilde{\alpha}(t) - \alpha_0(t) \\
\tilde{\theta}(t) - \theta_0(t)
\end{pmatrix} \right\| - \varepsilon_n^*(t).
\]

**Theorem 2.** Assume the conditions of Theorem 1 and (C5) hold and there exist iid random functions \(\{\phi_1(t)\}_{i=1}^{\infty}\) such that \(\sup_{t \in [l, u]} \|\sqrt{n}\{\tilde{\eta}(t) - \eta_0(t)\} - n^{-1/2} \sum_{i=1}^{n} \phi_1(t)\| \xrightarrow{P} 0\), and \(\{\phi_1(t), t \in [l, u]\}\) is Glivenko-Cantelli and Donsker. Then, \(\sqrt{n}\{\tilde{\alpha}(t) - \alpha_0(t)\} \sim \sqrt{n}\{\tilde{\theta}(t) - \theta_0(t)\}\) converges weakly to a zero-mean Gaussian process with covariance function \(\Sigma(s, t) = E\{W_1(s)W_1(t)\}\), where
\( W_i(t) = J(t)^{-1} \cdot [A_i \{ \alpha_0(t), \theta_0(t), \eta_0(t), t \} - H(t) \phi_i(t)] \), \( i = 1, \ldots, n \). Suppose there exists statistics \( \hat{\phi}_i(t) \), satisfying \( \sup_{t \in [l, u]} \| \hat{\phi}_i(t) - \phi_i(t) \| \rightarrow 0 \) and let

\[
\hat{W}_i(t) = \hat{J}(t)^{-1} \cdot [A_i \{ \hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t), t \} - \hat{H}(t) \hat{\phi}_i(t)], \quad \hat{\Sigma}(s, t) = n^{-1} \sum_{i=1}^{n} \hat{W}_i(t)\hat{W}_i(t)^T.
\]

Then \( \sup_{s \in [l, u]} \| \hat{\Sigma}(s, t) - \Sigma(s, t) \| \rightarrow 0 \), in probability.

**Proof of Theorem 2:** Simple algebra shows that

\[
0 = n^{-1/2} U \{ \hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t), t \}
= n^{-1/2} U \{ \alpha_0(t), \theta_0(t), \eta_0(t), t \} + n^{-1/2} \{ U \{ \hat{\alpha}(t), \hat{\theta}(t), \hat{\eta}(t), t \} - U \{ \alpha_0(t), \theta_0(t), \eta_0(t), t \} \}
= n^{-1/2} U \{ \alpha_0(t), \theta_0(t), \eta_0(t), t \}
+ n^{-1/2} \sum_{i=1}^{n} \{ K_i(\hat{\alpha}, \hat{\theta}, \hat{\eta}, \alpha_0, \theta_0, \eta_0, t) - K_i(\alpha_0, \theta_0, \eta_0, \alpha_0, \theta_0, \eta_0, t) \}
- n^{-1/2} \sum_{i=1}^{n} V_i(\hat{\alpha}(t), \hat{\theta}(t), t) D_i(\hat{\alpha}(t), \hat{\theta}(t), t) I(Y_i > t)
\cdot \left[ \Psi \{ \hat{\alpha}(t), \hat{\theta}(t)^T Z_i, \hat{\eta}(t)^T Z_i \} - \Psi \{ \alpha_0(t), \theta_0(t)^T Z_i, \eta_0(t)^T Z_i \} \right]
= U_0(t) + U_1(t) - U_2(t)
\]

Let \( \tilde{D}_i(\alpha(t), \theta(t), \eta(t)) = \partial \{ V_i(\alpha(t), \theta(t), t) D_i(\alpha(t), \theta(t), \eta(t)) \} / \partial \{ \alpha(t) \theta(t) \eta(t) \} \). From Taylor expansion of \( V_i(\alpha(t), \theta(t), t) D_i(\alpha(t), \theta(t), \eta(t)) \) at \( \{ \alpha_0(t), \theta_0(t), \eta_0(t) \} \),

\[
U_1(t) = \frac{1}{n} \sum_{i=1}^{n} \tilde{D}_i(\alpha_0(t), \theta_0(t), \eta_0(t)) \left[ I(X_i > t) - I(Y_i > t) \Psi \{ \alpha_0(t), \theta_0(t)^T Z_i, \eta_0(t)^T Z_i \} \right]
\cdot \sqrt{n} \left\{ \hat{\alpha}(t) - \alpha_0(t), \hat{\theta}(t) - \theta_0(t), \hat{\eta}(t) - \eta_0(t) \right\}^T + \nu_n(t).
\]

Uniform consistency of \( \hat{\alpha}, \hat{\theta} \) and \( \hat{\eta} \), coupled with (C5), imply \( \sup_{t \in [l, u]} \| \nu_n(t) \| \rightarrow 0 \), and hence

\[
\left\{ \tilde{D}_i(\alpha_0(t), \theta_0(t), \eta_0(t)) \left[ I(X_i > t) - I(Y_i > t) \Psi \{ \alpha_0(t), \theta_0(t)^T Z_i, \eta_0(t)^T Z_i \} \right] \right\}, \quad t \in [l, u]
\]

is Glivenko-Cantelli. Thus \( U_1(t) \) can be written as \( \tilde{v}_n(t) \sqrt{n} \left\{ \hat{\alpha}(t) - \alpha_0(t), \hat{\theta}(t) - \theta_0(t), \hat{\eta}(t) - \eta_0(t) \right\}^T + \nu_n(t) \), where \( \sup_{t \in [l, u]} \| \tilde{v}_n(t) \| \rightarrow 0 \).

Next, by applying Taylor expansion and uniform law of large numbers, \( U_2(t) = \{ J(t) + \nu_n(t) \cdot \sqrt{n} \left( \hat{\alpha}(t) - \alpha_0(t) \right) + \left[ H(t) + \nu_n(t) \cdot \sqrt{n} \left( \hat{\eta}(t) - \eta_0(t) \right) \right] \left[ \frac{1}{n} \sum_{i=1}^{n} \tilde{D}_i(\alpha_0(t), \theta_0(t), \eta_0(t)) \left[ I(X_i > t) - I(Y_i > t) \Psi \{ \alpha_0(t), \theta_0(t)^T Z_i, \eta_0(t)^T Z_i \} \right] \right] \}
\]

uniformly converge to 0 in probability for \( t \in [l, u] \). The validity of uniform law of large numbers follows from \( G^* \) and \( G^{**} \) being Glivenko-Cantelli. Thus,

\[
\sqrt{n} \left( \hat{\alpha}(t) - \alpha_0(t) \right) = J(t)^{-1} \left[ n^{-1/2} U \{ \alpha_0(t), \theta_0(t), \eta_0(t), t \} - H(t) \sqrt{n} \left( \hat{\eta}(t) - \eta_0(t) \right) \right] + \tau_n(t)
\]

21
where \( \sup_{t \in [l,u]} \| r_n(t) \| \xrightarrow{p} 0 \). Weak convergence follows since \( \{ A_1 \{ a_0(t), \theta_0(t), \eta_0(t), t \} : t \in [l, u] \} \), is a subclass of \( \mathcal{G} \), and \( \{ \phi_1(t) : t \in [l, u] \} \) are Donsker classes.

Uniform consistency of \( \hat{J}(t) \) and \( \hat{H}(t) \) for \( J(t) \) and \( H(t) \) follows from Glivenko-Cantelli property of \( \mathcal{G}^* \) and \( \mathcal{G}^{**} \) and uniform consistency of \( \hat{\alpha} \), \( \hat{\theta} \) and \( \hat{\eta} \). Since \( \{ A_1 \{ a_0(t), \theta_0(t), \eta_0(t), t \} , t \in [l, u] \} \) and \( \{ \phi_1(t) , t \in [l, u] \} \) are Glivenko-Cantelli, and \( J(t)^{-1} \) and \( H(t) \) are bounded for \( t \in [l, u] \), \( \{ W_1(t), t \in [l, u] \} \) is also Glivenko-Cantelli. By Slutsky's theorem and uniform law of large numbers, \( \hat{\Sigma}(s, t) \) converges to \( \Sigma(s, t) \) uniformly.
Table 1.

*Biases, empirical and model based variances, and coverage probabilities for $\hat{\theta}_2(t)$.*

<table>
<thead>
<tr>
<th>$n$</th>
<th>$t$</th>
<th>Bias</th>
<th>EmpVar</th>
<th>ModVar</th>
<th>CovProb</th>
<th>Bias</th>
<th>EmpVar</th>
<th>ModVar</th>
<th>CovProb</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.22</td>
<td>0.001</td>
<td>0.129</td>
<td>0.132</td>
<td>0.943</td>
<td>0.022</td>
<td>0.119</td>
<td>0.129</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>0.47</td>
<td>0.009</td>
<td>0.067</td>
<td>0.075</td>
<td>0.950</td>
<td>0.030</td>
<td>0.068</td>
<td>0.071</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>0.72</td>
<td>0.013</td>
<td>0.059</td>
<td>0.066</td>
<td>0.946</td>
<td>0.028</td>
<td>0.061</td>
<td>0.063</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td>0.97</td>
<td>0.004</td>
<td>0.066</td>
<td>0.067</td>
<td>0.944</td>
<td>0.027</td>
<td>0.062</td>
<td>0.064</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td>1.22</td>
<td>0.002</td>
<td>0.070</td>
<td>0.076</td>
<td>0.945</td>
<td>0.015</td>
<td>0.066</td>
<td>0.074</td>
<td>0.963</td>
</tr>
<tr>
<td>500</td>
<td>0.22</td>
<td>0.001</td>
<td>0.059</td>
<td>0.061</td>
<td>0.950</td>
<td>0.013</td>
<td>0.056</td>
<td>0.061</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td>0.47</td>
<td>0.012</td>
<td>0.034</td>
<td>0.039</td>
<td>0.951</td>
<td>0.030</td>
<td>0.033</td>
<td>0.037</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td>0.72</td>
<td>0.012</td>
<td>0.029</td>
<td>0.034</td>
<td>0.962</td>
<td>0.033</td>
<td>0.034</td>
<td>0.035</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>0.97</td>
<td>0.000</td>
<td>0.029</td>
<td>0.037</td>
<td>0.957</td>
<td>0.029</td>
<td>0.033</td>
<td>0.035</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td>1.22</td>
<td>0.007</td>
<td>0.033</td>
<td>0.042</td>
<td>0.944</td>
<td>0.017</td>
<td>0.037</td>
<td>0.038</td>
<td>0.932</td>
</tr>
</tbody>
</table>
Table 2.
Bias and empirical and model based variances for estimator of $\theta_{20}(t) = \theta_2$ and rejection rates for nominal 0.05 level tests of $H_0: \theta_{20}(t) = 1$ and $\hat{H}_0: \theta_{20}(t) = 0$.

<table>
<thead>
<tr>
<th>$(\theta_2, \eta_2)$</th>
<th>$n$</th>
<th>Bias</th>
<th>EmpVar</th>
<th>ModVar</th>
<th>$T_{12}$</th>
<th>$T_{13}$</th>
<th>$T_2$</th>
<th>$T_{12}$</th>
<th>$T_{13}$</th>
<th>$T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,0)</td>
<td>250</td>
<td>0.029</td>
<td>0.040</td>
<td>0.039</td>
<td>0.051</td>
<td>0.051</td>
<td>0.055</td>
<td>0.991</td>
<td>0.984</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.013</td>
<td>0.020</td>
<td>0.020</td>
<td>0.058</td>
<td>0.049</td>
<td>0.058</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>(1,0.2)</td>
<td>250</td>
<td>0.034</td>
<td>0.043</td>
<td>0.041</td>
<td>0.052</td>
<td>0.044</td>
<td>0.060</td>
<td>0.988</td>
<td>0.978</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.039</td>
<td>0.021</td>
<td>0.021</td>
<td>0.052</td>
<td>0.049</td>
<td>0.060</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 3.
Test statistics ($p$ values) for nonparametric tests of $H_0: \theta_{10}(t) = 0$.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$T_{13}$</th>
<th>$T_{14}$</th>
<th>$T_{15}$</th>
<th>$T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFV vs NFV</td>
<td>3.370 (0.338)</td>
<td>5.324 (0.256)</td>
<td>3.389 (0.640)</td>
<td>4.633 (0.031)</td>
</tr>
<tr>
<td>EFV+NFV vs NFV</td>
<td>10.263 (0.016)</td>
<td>11.091 (0.026)</td>
<td>10.321 (0.067)</td>
<td>13.694 (&lt;0.001)</td>
</tr>
<tr>
<td>New3TC</td>
<td>29.537 (&lt;0.001)</td>
<td>30.731 (&lt;0.001)</td>
<td>36.312 (&lt;0.001)</td>
<td>33.205 (&lt;0.001)</td>
</tr>
<tr>
<td>Baseline RNA</td>
<td>26.405 (&lt;0.001)</td>
<td>29.076 (&lt;0.001)</td>
<td>30.175 (&lt;0.001)</td>
<td>52.148 (&lt;0.001)</td>
</tr>
</tbody>
</table>

Note: $T_{1j}$: $T_1$ based on $j$ time points.

Table 4.
Estimated coefficients, standard errors and $p$ values by naively fitting a propotional hazards model to $T_1$ and by fitting $\theta_{10}(t) = \theta_{10}$ using $\hat{\theta}_{10}(t)$.

<table>
<thead>
<tr>
<th></th>
<th>Naive</th>
<th></th>
<th></th>
<th>New</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>SE</td>
<td>$p$ value</td>
<td>Est</td>
<td>SE</td>
<td>$p$ value</td>
</tr>
<tr>
<td>EFV vs NFV</td>
<td>-0.416</td>
<td>0.222</td>
<td>0.061</td>
<td>-0.487</td>
<td>0.226</td>
<td>0.031</td>
</tr>
<tr>
<td>EFV+NFV vs NFV</td>
<td>-1.339</td>
<td>0.284</td>
<td>&lt;0.001</td>
<td>-1.130</td>
<td>0.305</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>New3TC</td>
<td>-1.389</td>
<td>0.258</td>
<td>&lt;0.001</td>
<td>-1.488</td>
<td>0.258</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Baseline RNA</td>
<td>0.730</td>
<td>0.134</td>
<td>&lt;0.001</td>
<td>0.834</td>
<td>0.116</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Table 5.

Goodness-of-fit tests (p values) for $\theta_{i0}(t) = \theta_{i0}$.

<table>
<thead>
<tr>
<th></th>
<th>$T_{13}^*$</th>
<th>$T_{14}^*$</th>
<th>$T_{21}^*$</th>
<th>$T_{22}^*$</th>
<th>$T_{23}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFV vs NFV</td>
<td>1.945 (0.582)</td>
<td>3.595 (0.463)</td>
<td>0.218 (0.641)</td>
<td>0.296 (0.586)</td>
<td>0.298 (0.585)</td>
</tr>
<tr>
<td>EFV+NFV vs NFV</td>
<td>2.690 (0.442)</td>
<td>2.946 (0.567)</td>
<td>0.541 (0.462)</td>
<td>0.651 (0.420)</td>
<td>0.628 (0.428)</td>
</tr>
<tr>
<td>New3TC</td>
<td>0.553 (0.907)</td>
<td>1.069 (0.898)</td>
<td>0.0002 (0.989)</td>
<td>0.097 (0.755)</td>
<td>0.165 (0.685)</td>
</tr>
<tr>
<td>Baseline RNA</td>
<td>4.729 (0.193)</td>
<td>0.478 (0.976)</td>
<td>0.211 (0.645)</td>
<td>0.245 (0.621)</td>
<td>0.290 (0.590)</td>
</tr>
</tbody>
</table>
Figure 1. Point estimates and 0.95 pointwise confidence intervals for time-varying copula parameter and covariate effects. The ragged solid lines are the point estimates, the dashed lines are the 0.95 pointwise confidence intervals, the dotted lines are a lowess smoothing curve from the estimated parameters, and the horizontal solid lines are the fitted covariate effects from constant models. (a) copula parameter, (b) EFV vs NFV, (c) EFV+NFV vs NFV, (d) New3TC, (e) log_{10} baseline RNA.