This is the author version of an article published as:


**Access to the published version:**

[http://dx.doi.org/10.1016/j.insmatheco.2009.11.001](http://dx.doi.org/10.1016/j.insmatheco.2009.11.001)

**Copyright:** Elsevier 2009
Applying Copula Models to Individual Claim Loss Reserving Methods*

XiaoBing ZHAO$^{1,2}$, Xian ZHOU$^3$

$^1$School of Science, Jiangnan University, Wuxi, Jiangsu Province, 214122, PRC
$^2$School of Finance and Statistics, East China Normal University, Shanghai, PRC
$^3$Department of Actuarial Studies, Macquarie University, NSW, Australia

Abstract. The estimation of loss reserves for incurred but not reported (IBNR) claims presents an important task for insurance companies to predict their liabilities. Recently, individual claim loss models have attracted a great deal of interest in the actuarial literature, which overcome some shortcomings of aggregated claim loss models. The dependence of the event times with the delays is a crucial issue for estimating the claim loss reserving. In this paper, we propose to use semi-competing risks copula and semi-survival copula models to fit the dependence structure of the event times with delays in the individual claim loss model. A nonstandard two-step procedure is applied to our setting in which the associate parameter and one margin are estimated based on an ad hoc estimator of the other margin. The asymptotic properties of the estimators are established as well. A simulation study is carried out to evaluate the performance of the proposed methods.

Key words: IBNR claim, individual claim loss model, event and delay times, Clayton copula, semi-competing risks, semi-survival copula.

*This work was partially supported by the NSFC, under Grant No.10871084, PIRTJiangnan, and Macquarie University Research Development Grant.
1. Introduction

The incurred but not reported (IBNR) claim is a classic problem extensively studied in the actuarial literature. The IBNR claims include IBNYR (incurred but not yet reported) claims whose occurrence in the insured exposure interval is not known until some later date due to random reporting delay and IBNER (Incurred but not enough reported) claims whose occurrence is always known but with incomplete cost development (cf. Jewell, 1989). The estimation of loss reserves for IBNR claims presents an important and challenging task for insurers to correctly predict their liabilities.

The standard approach to reserving for IBNR claims usually applies one of the grouped (or aggregate) data models. Techniques such as chain ladder method or separation method can be used to predict the unknown (not yet reported) claims. To date, however, more sophisticated statistical models and methodologies, which have been extensively applied in many practices, do not seem to have been adequately applied to IBNR problems. Exceptions include Spreeuw and Goovaerts (1998), who proposed a discrete hazard rate model for reporting time to predict the future claim numbers, which extends the work of Kaminsky (1987). In addition, a randomly truncated data model for reserving IBNR claims was proposed by Herbst (1999), in which the joint distribution of the claim severity and delay is estimated by using a nonparametric method based on left-truncated delays (by the event time).

The statistical analyses in the above mentioned papers are based on the grouped or aggregate data, using the so-called “run-off triangle”, which have been illustrated to have some drawbacks by many authors, including Doray (1994), Hesselager and Witting (1988), Taylor et al. (2008) and Larsen (2007). The first shortcoming is the assumption of the independence in the reported claim numbers between early and late development years (cf. Hesselager and Witting, 1988), which is often violated in practice. The second is that the information of individual claims is not fully utilized in the traditional claims reserving methods (Jewell, 1989, 1990), since the “triangulation of data” discards a great deal of information. Some, or even much, of such information may be valuable to the forecast required (Taylor et al., 2008). Another shortcoming is the assumption of identically distributed claim frequency, severity and delays, which are often unrealistic in practice due to seasonal effects and changes in portfolio size, mix of business, claim types, and empirical claim distributions (Larsen, 2007). If these shortcomings of conventional methods are ignored, the prediction of the claim loss reserving could be inaccurate and misleading.
Recently, the individual claim loss model has attracted a great deal of interest in the literature. It was first proposed by Norberg (1986, 1993) and Jewell (1989, 1990) (cf. also Hachemeister, 1980, and Haastrup and Arjas, 1996) in an attempt to lay down a comprehensive architecture of claims, and subsequently investigated extensively by Taylor et al. (2008) and Larsen (2007). This individual claim loss model has a number of advantages over conventional models based on grouped or aggregate claims – referred to as the aggregate claim model by Taylor et al. (2008).

Meanwhile, the copula construction has been popular in the financial and actuarial literature, which turns out to be very useful to model the dependence in finance, actuarial science and survival analysis. There are several possibilities to work with copulas. First, one can assume parametric models for both copulas and marginal distributions. The model parameters can be estimated by maximum likelihood or inference function for margins; see Oakes (1982), Romano (2002) and Joe (2005) for more details of these methods. A second possibility is to consider nonparametric models for both copulas and marginal distributions. Deheuvels (1979) proposed a method based on the multivariate empirical distribution. Gijbels and Międlarczyk (1990) and Bouezmarni and Rombouts (2009) used a kernel method to estimate a bivariate copula. More recently, Chen and Huang (2007) proposed a bivariate estimator based on the local linear estimator, and Morettin et al. (2006) proposed an empirical copulas wavelet estimator. A third possibility is a semiparametric approach, which combines a parametric model for copulas and a nonparametric model for marginal distributions. This semiparametric model was initially developed by Oakes (1986), Genest et al. (1995) and Genest and Rivest (1993), and further researched by Shih and Louis (1995), Glidden (2000) and He and Lawless (2003) with extensions to censored observations and/or proportional hazards structure for marginal distributions. In addition, Tsukahara (2007) proposed pseudo Z-estimator and Chen et al. (2006) utilized the sieve maximum likelihood.

In this paper, we note that the reported times are censored by the endpoint of the reported data, say \( b \), which implies that the delay is censored by the variable \( b - t \), where \( t \) is the corresponding event time. There have been many discussions in the literature on censoring schemes involving bivariate variables and copula models. Apart from the bivariate right censoring (cf. Shih and Louis, 1995; Wang and Wells, 2000), the existing censoring schemes in copula models can be classified into three categories:

(i) One variable is (independently or dependently) censored by the other variable, such as dependent censoring in Rivest and Wells (2001), Braekers and Veraverbeke (2005) and Li et al. (2007).

(ii) Only one component of a bi-variable is subject to censoring by a third party, which is discussed in Akritas (1994) and Denuit et al. (2004).

(iii) Semi-competing risk model, in which one event censors the other, but not vice versa, see for example Peng and Fine (2007) and Chaieb-Lakhal et al. (2008). This censoring scheme is suitable for our setting with the delay \( w \) being censored by \( b - t \).

Since the dependence of the event times on the delays is of paramount importance in IBNR loss reserving, and motivated by the work of Denuit et al. (2004), Taylor et al. (2008) and Larsen (2007), we propose in this paper a new individual claim loss model to account for such dependence. In our model, the event time and the corresponding delay are associated by a parametric copula under the semi-competing risk model. In addition, the arrival process of the claim counts is modeled by a non-stationary Poisson process with an unspecified baseline intensity function and covariates; the hazard rate function of the delays is modeled by a proportional hazards structure with a non-parametric baseline distribution. We also propose a nonstandard two-step procedure to estimate the parametric and nonparametric components of our model and establish the asymptotic properties of the estimators. This semiparametric model together with the proposed methodology is more flexible than the previous models for IBNR problems in the literature, and is expected to produce more effective and accurate predictions for the claim loss reserving.

The rest of the paper is organized as follows. In Section 2 we review existing individual claim models and specify our model. Section 3 discusses different censoring and truncation mechanisms. A nonstandard two-step procedure to estimate the parameters is proposed in Section 4 and the asymptotic properties of the estimators are provided in Section 5. Section 6 reports some simulation results, and Section 7 concludes.
2. Model specifications

Individual claim loss models have been investigated (quite properly) under very general frameworks. The statistical analysis of the prediction of IBNR events and delays are typically based on the following two assumptions (cf. Jewell, 1989):

\( (A_1) \) The events (claims) of interest are generated from a homogeneous Poisson process with rate \( \lambda \) (claim/year) over some fixed interval \((0, a]\) (exposure interval). There is an unknown number \( \tilde{n} = \tilde{n}(a) \) of events at unknown occurrence epochs (accident dates) \( \tilde{x}_1, \ldots, \tilde{x}_{\tilde{n}} \). It follows that \( \tilde{n} \) has a Poisson distribution with mean \( \lambda a \), and given \( \tilde{n} = n \), the epochs \( \tilde{x}_1, \ldots, \tilde{x}_n \) (not ordered) are independent random variables uniformly distributed over \((0, a]\).

\( (A_2) \) Each event \( j \) is associated with a positive random waiting time (reporting delay), say, \( \tilde{w}_j > 0 \), so that the observation epochs (reporting data) are given by \( \tilde{y}_j = \tilde{x}_j + \tilde{w}_j \), \( j = 1, 2, \ldots, n \). Also assume that the delays \{\( \tilde{w}_j \)\} are independent and identically distributed (i.i.d) random variables with a common density \( f_w(\cdot|\theta) \) and cumulative distribution function (cdf) \( F_w(\cdot|\theta) \), where \( \theta \) is an unknown parameter (vector). Both \( f_w \) and \( F_w \) have supports in \([0, \infty)\).

The individual claim loss model proposed by Jewell (1989, 1990) has advantages over the aggregate claim model in many aspects. It does not, however, overcome some major shortcomings of the aggregate claim model for the prediction of the claim loss, such as the assumption of a stationary Poisson process for the arrivals of claims, which rarely holds in reality. As discussed in Taylor et al. (2008) and Larsen (2007), claims are often time-dependent and associated with covariates. Hence a non-stationary Poisson process would be more appropriate to allow for heterogeneity over time. Furthermore, the assumption of parameterized i.i.d. delays is also a strong condition that is hardly matched in practice. Thus a natural extension of the work of Jewell (1990) is to take account of the individual claim loss model proposed by Taylor et al. (2008) and Larsen (2007), which is defined through assumptions \((B_1)-(B_3)\) below.

For subject (policy) \( i \), let \( N_i(t) \) denote the number of recurrent events (claims) occurred up to time \( t \), \( X_{i1} \) a vector of time-independent or time-dependent covariates, and \( a \) represent the common terminating time for observing the event process \( N_i(t) \). We further denote by \( m_i \) the number of claims up to time \( a \) and \( t_{i1}, t_{i2}, \ldots, t_{im_i} \) the observed event times (occurrence epochs) for subject \( i \). For ease of notation, we use \( m_i \) and \( t_{ij}, j = 1, 2, \ldots, m_i, i = 1, 2, \ldots, n \), to denote either random variables or their
realized values. The first assumption for our proposed model is:

(B1) Given $X_{i1} = x_{i1}$, the recurrent event process $N_i(\cdot)$ follows a non-stationary Poisson process with a multiplicative intensity (rate) function

$$\lambda_i(t|x_{i1}) = \lambda_0(t) \exp(x_{i1}^T \beta), \quad 0 \leq t \leq a, \quad (2.1)$$

where $\beta$ is a $p \times 1$ vector of parameters and the baseline intensity function $\lambda_0(t)$ is a continuous function.

For subject $i$, given the number $m_i$ of claims by time $t$, we define $\Lambda_i(t) = \int_0^t \lambda_i(s) ds$, where $\lambda_i(t) = \lambda_i(t|x_{i1})$ is defined in (2.1), and $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$.

Let $0 < t_{i1} < t_{i2} < \cdots < t_{imi} < a$. Similar to Theorem 2.3.1 in Ross (1996), we can show that the joint conditional density function of $t_{i1}, \ldots, t_{imi}$ given $N_i(a) = m_i$ is independent of covariates, given by

$$f(t_{i1}, \ldots, t_{imi}) = \frac{(m_i)!}{[\Lambda_i(a)]^{m_i}} \prod_{j=1}^{m_i} \lambda_i(t_{ij}). \quad (2.2)$$

(2.2) implies that, given $(x_{i1}, m_i)$, the observed occurrence epochs $\{t_{i1}, t_{i2}, \ldots, t_{imi}\}$ are the order statistics of i.i.d. random variables with density

$$\pi(t) = \frac{\lambda_0(t) \exp(x_{i1}^T \beta)}{\Lambda_0(a) \exp(x_{i1}^T \beta)} = \frac{\lambda_0(t)}{\Lambda_0(a)}, \quad 0 \leq t \leq a. \quad (2.3)$$

For subject $i$, each event occurred at the event epoch $t_{ij}$ is also associated with a positive waiting time to report, say $w_{ij}$. The analysis of Hesselager and Witting (1988) indicates a negative correlation in claim numbers between early and late development years (see also discussions of Larsen, 2007 and Taylor et al., 2008). Hence the assumption of i.i.d. delays is violated. In this paper, an alternative semiparametric model of delays is proposed to accommodate various covariates, which is given in the second assumption of our model:

(B2) Given a covariate vector $x_{i2}$, which may be time-independent or time-dependent, the delays $w_{ij}$ have a semiparametric structure with a hazard rate function

$$h_i(w|x_{i2}) = h_0(w) \exp(x_{i2}^T \alpha), \quad w > 0, \quad (2.4)$$

where $\alpha$ is a $p \times 1$ vector of parameters and $h_0(\cdot)$ is a continuous baseline hazard function.
The reporting time of the claim at time \( t_{ij} \) is \( y_{ij} = t_{ij} + w_{ij} \). We further assume that all reported events are in observation interval \((0, b]\). This leads to an observed number of reported events, say \( r_i(b) \), such that \( y_{ij} = t_{ij} + w_{ij} \leq b \), \( j = 1, 2, \ldots, m_i \). The remaining \( u_i = m_i(a) - r_i(b) \) events are unreported events with \( y_{ij} = t_{ij} + w_{ij} > b \). The model in (2.4) has a semiparametric structure, which is more flexible than the parametric model proposed by Taylor et al (2008) and Jewell (1990).

The relationship between the event time and the delay is assumed to be independent in Jewell (1990). Given the delay distribution, the conditional distribution of the claims was proposed by Hesselager and Witting (1988) to depict the relationship between correlated delays. Recently, copula has been extensively used as a tool for modeling dependence structures in different fields of applications: finance, insurance, risk theory, environmental studies, etc. In this paper, we characterize the dependence structure between the event time and the delay by a parameterized copula with the margins of the event time and the delay defined in (2.3) and (2.4) respectively. Thus the third assumption of our model is:

\[ (B_3) \] Given a bivariate survival function \( S(\cdot, \cdot) \) with univariate marginal survival functions \( S_T(\cdot) \) and \( S_W(\cdot) \), there exists a copula \( C(\cdot, \cdot) \) such that for all \( (t, w) \in \mathbb{R}^2 \), the joint survival function \( S(\cdot, \cdot) \) of \( (T, W) \) can be represented by:

\[ S(t, w) = C(S_T(t), S_W(w)), \quad (t, w) \in \mathbb{R}^2. \] (2.5)

We then propose a nonstandard two-stage estimation procedure to estimate the copula parameters using ad hoc estimators of the marginal distributions.

As an example of model (2.5), Figures 1 and 2 in Hesselager and Witting (1988) illustrated a negative correlation between the claim numbers in early and late development years, in the sense that a smaller number of claims observed in early development years is associated with a larger number of claims in late development years (as compared to the expected pattern), and vice versa. This provides the evidence of the dependence between the claim arrivals and the delays.

The proposed model is distinct from the existing models for IBNR claim loss reserving in two aspects: First, all data are utilized to construct the estimates under semi-competing risks model, whereas the existing models use only uncensored reported data (cf. Jewell, 1990, Petere, 2006 and Petere and Kollo, 2006). Secondly, the censoring mechanism in our model is based on the dependent relationship between the event time and the delay through a copula and the margin of the delay has a PH structure (2.4),
which is more flexible than that of Herbst (1999) with a nonparametric joint distribution of dependent variables.

This paper will investigate the proposed model (2.1) together with (2.4) and (2.5). Under Assumptions \((B_1) - (B_3)\), we can draw statistical inference on the parameters of the baseline and the coefficients of the covariates from the data, and then determine the prediction of the claim loss.

3. Censoring, truncation and copula

Several estimating procedures for the copula parameter and the ensuing marginals have been proposed under different incomplete-data schemes. Examples include bivariate censored data (Shih and Louis, 1995), semi-competing risks data (Fine et al., 2001) and semi-competing risks data subject to truncation (Chaieb-Lakhal et al., 2006). In order to construct the two-stage likelihood for individual claim loss reserving model, two incomplete-data schemes are discussed below:

- **Case I**: The observable pairs \((b - t_{ij}, w_{ij})\) follow a semi-competing risk model, in which \(w_{ij}\) may be censored by \(b - t_{ij}\). This setting has been considered in the literature of survival analysis by many authors such as Fine et al. (2001), Wang (2003), Peng and Fine (2007) and Hsieh and Wang (2008). Under this setting, however, the censoring mechanism of the event time and the delay has not been considered so far, which corresponds to the case of IBNER claims.

- **Case II**: The study subjects (i.e., \(t_{ij}\)) consist of all occurrence epochs (accident dates or event times) whose secondary events (the delays \(w_{ij}\)) occur in the time interval \([0, b - t_{ij}]\). Once the secondary event \(w_{ij}\) is observed, the time \(t_{ij}\) of the primary event is also known. The data are right-truncated (Turnbull, 1976 and Cui, 1999) because only the pairs \((t_{ij}, w_{ij})\) with \(w_{ij} \leq b - t_{ij}\) are observed; otherwise if \(w_{ij} > b - t_{ij}\), nothing is observed on the pair. This setting corresponds to the case of IBNYR claims. The truncation considered in this paper, however, is dependent on the event time, which is more realistic than the independent truncation in Cui (1999). This dependent truncation allows us to utilize the semi-competing risk model, but with censoring replaced by truncation.

In insurance mathematics, copulas have become a common tool since Frees and Valdez (1998). Usually, claim sizes and expenses are of the main interest in copula models. In this paper we accentuate to the event time and its delay.
Let \((T, W)\) denote the paired variables representing the event (claim) time and the corresponding delay with marginal survival functions \((S_T(\cdot), S_W(\cdot))\) and marginal densities \((f_T(\cdot), f_W(\cdot))\). If \((T, W)\) have a joint survival function \(S(t, w) = C_\phi(S_T(t), S_W(w))\) given by a copula \(C_\phi\), then \(S\) is said to be generated by \(C_\phi\), and the joint density of \((T, W)\) is given by

\[
    f(t, w) = \frac{\partial^2 C_\phi(S_T(t), S_W(w))}{\partial t \partial w}, \quad t \geq 0, \; w \geq 0.
\]

Let \(\phi : [0, +\infty] \to [0, 1]\) be a twice-differentiable strictly decreasing and convex function, which implies that \(\phi\) has a twice-differentiable inverse \(\phi^{-1}\). Every such function \(\phi\) generates a copula function

\[
    C_\phi(u, v) = \begin{cases} 
        \phi\{\phi^{-1}(u) + \phi^{-1}(v)\}, & \text{if } 0 \leq u, v \leq 1, \\
        0, & \text{otherwise}, 
    \end{cases}
\]

with \(0 \leq \phi \leq 1\), \(\phi(0) = 1\), \(\phi' < 0\) and \(\phi'' > 0\). A copula \(C_\phi\) of the form (3.1) is referred to as an Archimedean copula. The function \(\phi\) is called the generator of the copula. Only those \(\phi\) satisfying \(\lim_{t \to +\infty} \phi(t) = 0^+\) are used in this work.

There is a large literature to discuss how to choose a copula for a given dataset, such as Genest and Rivest (1993), which developed a procedure for identifying an appropriate copula. Just as in Pettere and Kollo (2006), we present the Clayton copula as an example for our statistical inference in this paper. The Clayton copula is the most frequently used and best known Archimedean copula in empirical application (Cherubini, et al., 2004). It has a number of advantages including mathematical tractability and some nice properties such as a constant association for time-dependence (the driver of the dependence; cf. Chaieb-Lakhal, 2006, 2008) under a measure of Oakes’s cross-ratio function (Oakes, 1989) and adaptation to a random effects model with alternative but equivalent parameterization (cf. Glidden, 2000). Furthermore, the Clayton copula is the only absolutely continuous copula that is preserved under truncation (cf. Oakes, 2005). Consequently, if the \(C(\cdot, \cdot)\) in (2.5) is a Clayton copula, then conditional on \(T > t\), the joint distribution of \(b - T\) and \(W\) is also a Clayton copula (with the same parameter if it is assumed to be fixed).

A bivariate survival function \(C_\theta(\cdot, \cdot)\) is said to be a \emph{Clayton copula} (Clayton, 1978) if it has the form

\[
    C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad \theta > 0.
\]

A Clayton copula is an Archimedean copula with generator \(\phi_\theta = (1 + t)^{-1/\theta}\). \(T\) and \(W\) are positively associated when \(\theta > 0\) and independent when \(\theta \to 0\).
4. Parameter estimation

4.1 Outline of nonparametric estimating equation

Semi-competing risk models have been extensively investigated in survival analysis, and can be used to model the relationship between the claims and their delays. In this subsection, we give a summary of the nonparametric estimation methodology of Peng and Fine (2007) for semi-competing risk models.

Let \( T_1 \) be the time to a non-terminating event, \( T_2 \) the time to a terminating event that may dependently censor \( T_1 \), and \( Z^0 \) a \( p \times 1 \) covariate vector. Define \( Y = T_1 \wedge T_2 \) and \( \delta = I(T_1 \leq T_2) \), where \( \wedge \) is the minimum operator and \( I(\cdot) \) is the indicator function. The observed data consist of \( n \) replicates of \((Y, \delta, Z^0)\), denoted by \( \{Y_i, \delta_i, Z^0_i\}, i = 1, \ldots, n \).

The data in Peng and Fine (2007) and Fine et al. (2001) allow another censoring variable independent of \((T_1, T_2)\), which are referred to as the semi-competing risks data with informative censoring. If we take this censoring variable as \(+\infty\), then such data have the same form as \( \{Y_i, \delta_i, Z^0_i\} \). With some appropriate modifications, the method of Peng and Fine (2007) can be applied to the data in our setting. The estimation methods are outlined as follow.

Assume that the hazard function for \( T_1 \) conditional on \( Z^0 \) satisfies

\[
\lambda(t|Z^0) = \lambda_0(t) \exp(\beta^T Z^0), \tag{4.1}
\]

where \( \lambda_0(t) \) is an unspecified baseline hazard function and \( \beta \) is a \( p \times 1 \) coefficient vector. Denote the cdf of \( T_1 \) by \( F(t|Z^0) \). Then \( \lambda(t|Z^0) = dF(t|Z^0)/[1 - F(t|Z^0)] \) by the definition of the hazard function. Model (4.1) implies

\[
\Pr(T_1 > t|Z^0) = \exp \{- \exp(\log \Lambda_0(t) + \beta^T Z^0)\} = g(\eta(t)^\top Z), \tag{4.2}
\]

where \( g(x) = \exp\{- \exp(x)\} \), \( \eta(t) = \{\log \Lambda_0(t), \beta^\top\} \) and \( Z = (1, Z^0)^\top \).

If the dependent censoring of \( T_1 \) by \( T_2 \) is ignored, then the naive partial likelihood estimator is biased. To overcome this, a joint model of \((T_1, T_2)\) is proposed by Peng and Fine (2007) for inference about the covariate effects on \( T_1 \) when \( T_1 \) is dependently censored by \( T_2 \). This is defined through a known time-dependent copula function \( C(s, t, \theta) \), where for fixed \( \theta \), \( C(\cdot, \cdot, \theta) \) satisfies the definition of a copula, and an unspecified time-varying parameter \( \theta = \theta(s, t) \), which is assumed to be cadlag (right-continuous with left limits). Let \( S_{T_i}(t|Z_i) \) be the survival function of \( T_i \) \((i = 1, 2)\). Assume that in the observable region

\[
\Pr(T_1 > s, T_2 > t|Z) = C\{S_{T_1}(s|Z), S_{T_2}(s|Z), \theta(s, t)\} \quad \text{for} \quad 0 \leq s \leq t, \tag{4.3}
\]
and for some known link function \( k(\cdot) \) and unknown function \( \vartheta = \vartheta(t) \),
\[
\Pr(T_2 > t | Z) = k\{\vartheta(t)^\top Z\}. \tag{4.4}
\]
Since \( T_1 \) is dependently censored by \( T_2 \), usual estimation methods do not work. Hence we need a simultaneous estimation of the covariate effects on \( T_1 \) and the dependence parameters for semi-competing risk models. Let \( \theta(t) = \theta(t, t) \). The following notations are needed to estimate \((\theta(t), \eta(t))\):
\[
A_i\{\theta(t), \eta(t), \vartheta(t), t\} = V_i D_i [I(Y_i > t) - I(T_{2i} > t)] \Psi_i,
\]
where \( V_i = V_i\{\theta(t), \eta(t), t\} \) is a scalar weight function, \( \Psi_i = \Psi_i(\theta(t), \eta(t), \vartheta(t)) = C\{S_{T_1}(t|Z_i), S_{T_2}(t|Z_i), \theta(t)\}/S_{T_2}(t|Z_i) \), and
\[
D_i = D_i\{\theta(t), \eta(t), \vartheta(t)\} = \frac{\partial \Psi_i(\theta(t), \eta(t), \vartheta(t))}{\partial(\theta(t), \eta(t))^\top}.
\]
Under models (4.2)–(4.4), we have \( E[I(Y_i > t)|I(T_{2i} > t), Z_i] = \Psi_i(\theta(t), \eta(t), \vartheta(t)) \) so that \( E[A_i\{\theta(t), \eta(t), \vartheta(t), t\}|I(T_{2i} > t), Z_i] = 0 \). This is a nonlinear binary regression model for the effects of \( Z_i \) on \( I\{\min(T_{1i}, T_{2i}) > t\} \) given \( T_{2i} > t \). Thus an estimating equation is proposed by Peng and Fine (2007) to estimate \((\theta(t), \eta(t))\) with \( \vartheta(t) \) replaced by its \textit{ad hoc} estimator \( \hat{\vartheta}(t) \):
\[
U\{\theta(t), \eta(t), \hat{\vartheta}(t), t\} = \sum_{i=1}^n A_i\{\theta(t), \eta(t), \hat{\vartheta}(t), t\}. \tag{4.5}
\]
The nonlinear estimating equation (4.5) jointly estimates \( \theta(t) \) and \( \eta(t) \) at each \( t \), adopting the “working independence” assumption across time (i.e., the observations from the same subject \( i \) are assumed to be independent). This greatly simplifies the computations, which would otherwise be rather complicated for estimating equations that combine information across time, especially because \( \Psi_i \) is highly nonlinear. It can be shown that the estimators of \( \theta(t) \) and \( \eta(t) \) are step functions, with jump only at observed failure (uncensored) points. Thus the estimating equation only needs to be solved at a finite number of time points (Peng and Fine, 2007).

In Subsections 4.2 and 4.3 next, we will investigate the parameter estimation of the proposed model under different censoring and truncation schemes. Since the observable pairs \((b - t_{ij}, w_{ij})\) are of main interest, we rewrite \( S_T(t) \) as the survival function of \( b - T \) in the sequel, that is,
\[
S_T(t) = \Pi(b - t) \quad \text{with} \quad \Pi(t) = \frac{\Lambda_0(t)}{\Lambda_0(a)} = \int_0^t \pi(s)ds \quad \text{and} \quad \pi(s) \text{ defined in (2.3)}.
\]
4.2 Semi-competing risks model

Case I censoring scheme (see page 8) has been studied on semi-competing risks data. Without covariates, Fine et al. (2001) developed inferences for a semiparametric model with dependence structure satisfying the gamma frailty model, which is the same as the Clayton copula, and unspecified marginal distributions. These inferences were extended to other parametric copulas by Wang (2003). More recently, a regression analysis based on semi-competing risks data has been investigated by Peng and Fine (2007), who incorporated covariates and formulated their effects on the survival function of the intermediate event via a functional regression model. To accommodate informative censoring, a time-dependent copula model is proposed in the observable region of the data that is more flexible than the standard parametric copula models for dependence between events. The methodology proposed by Hsieh and Wang (2008) is developed for discrete covariates under two types of assumptions in which separate copula models are assumed for each covariate group and then a flexible regression model is imposed on the progression time.

In the proposed models (2.3)–(2.5), it is of major interest to estimate the copula parameter \( \theta \), the marginal survival function \( S_T(t) \) of \( b - T \), the parameter \( \alpha \), and the unspecified baseline \( h_0(t) \) defined in (2.4). This can be carried out by using the estimation methods presented in Subsection 4.1. It implies that, based on an ad hoc estimator of \( S_T(t) \), we can estimate \( \alpha \) and \( h_0(t) \) simultaneously under the assumption of the survival function \( S_W(t|X_2) = g(\eta(t)^\top X_3) \), where \( g(x) = \exp(-\exp(x)) \), \( \eta(t) = [\log H_0(t), \alpha]^\top \) and \( X_3 = [1, X_2]^\top \). Then the model of Peng and Fine (2007) with a time-dependent copula reduces to a time-independent proportional hazards model proposed in this paper.

Define \( Y_{ij} = (b - T_{ij}) \wedge W_{ij} \) and \( \delta_{ij} = I(W_{ij} \leq b - T_{ij}) \). The observed data consist of \( m_i \) replicates of \((Y, \delta, X_2)\). With \( H_0(t) = \int_0^t h_0(s)ds \), model (2.4) implies

\[
\Pr(W > w|X_2) = S_W(w|X_2) = \exp\{-\exp(\log H_0(w) + \alpha^\top X_2)\}. \tag{4.6}
\]

With the semi-competing risks data, we link the joint distribution of \((T, W)\) to its marginals through a known time-dependent copula function \( C(u, v; \theta) \), where \( \theta = \theta(t, w) \) is an unspecified time-varying parameter, and is assumed to be cadlag. Assume that in the observable region \( 0 < w \leq b - t \),

\[
\Pr(b - T > t, W > w|X_2) = C\{S_T(t), S_W(w|X_2); \theta(t, w)\}, \quad 0 < w < b - t, \tag{4.7}
\]

where \( S_T(t) \) is the survival function of \( b - T \).
If the censoring time $b - T_{ij}$ is always observed, simultaneous estimators of $\alpha$, $h_0(t)$ and the copula parameter $\theta$ can be obtained by inserting an *ad hoc* estimator of $S_T(t)$ into the following estimating equation (cf. (4.5)):

$$\sum_{i=1}^{n} U_i \{ \theta(t), \eta(t), S_T(t), t \} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} A_{ij} \{ \theta(t), \eta(t), S_T(t), t \},$$

where $\eta(t) = (H_0(t), \alpha) \top$, $\theta(t) = \theta(t, t)$,

$$A_{ij} \{ \theta(t), \eta(t), S_T(t), t \} = V_{ij} D_{ij} \left[ I(Y_{ij} > t) - I(b - T_{ij} > t) \Psi_{ij} \right],$$

$V_{ij} = V_{ij} \{ \theta(t), \eta(t), t \}$ is a scalar weight function,

$$\Psi_{ij} = \Psi_{ij}(\theta(t), w, t) = C(S_T(t), S_{W_{ij}}(w), \theta(t))/S_T(t),$$

and

$$D_{ij} = D_{ij} \{ \theta(t), \eta(t), S_T(t) \} = \frac{\partial \Psi_{ij}(\theta(t), w, t)}{\partial (\theta(t), \eta(t)) \top}.$$

Under the model assumptions, $A_{ij} \{ \theta(t), \eta(t), S_T(t), t \}$ has zero mean conditional on $X_3$ and $I(b - T_{ij} > t)$ since $E[I(Y_{ij} > t)|b - T_{ij} > t, X_3] = \Psi_{ij} \{ \theta(t), \eta(t), S_T(t) \}$. Then the desired estimators can be obtained from (4.8). Obviously, the estimators of $\theta(t)$ and $\eta(t)$ are step functions, with jump only at observed failure (uncensored) points.

### 4.3 Semi-survival copula model

Subsection 4.2 discussed Case I of incomplete-data scheme, which accounts for censoring but not truncation in the sense that the variable $b - T_{ij}$ is assumed to be always observable; that is, the time of loss is known even if it has not yet been reported. We now consider Case II of incomplete-data scheme, which allows dependent truncation and has attracted growing interest in modeling survival data. In practical situations for IBNR claims, we often only observe pairs $(b - T_{ij}, W_{ij})$ satisfying $W_{ij} \leq b - T_{ij}$; in other words, a loss is not known until it is reported. In such a case, $b - T_{ij}$ is left-truncated (by $W_{ij}$) and $W_{ij}$ is right-truncated (by $b - T_{ij}$). Censoring and truncation are fundamentally different. Censored subjects are known to be in the sample, but their failure times are not available. Truncated individuals are completely missed by the data collection protocol. Models with dependent truncation have been investigated by many authors. For example, Peng and Fine (2006) proposed a nonparametric estimation with left-truncated semi-competing risks data, and Chaieb-Lakhal *et al.* (2006) generalized
the copula-graphic estimator of Zheng and Klein (1995) to truncated variables. In this paper, however, we focus on proportional hazards structure for one component. We present new statistical methods for modeling possible dependence between \(b - T_{ij}\) and \(W_{ij}\) through copula function when only pairs \((b - T_{ij}, W_{ij})\) such that \(W_{ij} \leq b - T_{ij}\) are observed, with one component assumed to be nonparametric and the other modeled by a proportional hazards structure.

It is convenient to write the joint distribution of \((b - T_{ij}, W_{ij})\) as

\[
R_{ij}(w, t) = \Pr\{W_{ij} \leq w, b - T_{ij} > t | b - T_{ij} > W_{ij}\} = \frac{1}{d_{ij}} \tilde{C}\{F_{W_{ij}}(w), S_T(t)\} \quad (4.9)
\]

for \(t \geq w\), where \(F_{W_{ij}}(w)\) and \(1 - S_T(t)\) are the cdf’s of \(W_{ij}\) and \(b - T_{ij}\) with densities \(f_{W_{ij}}(w)\) and \(f_T(t)\), respectively, \(\tilde{C}(\cdot, \cdot)\) is a copula function,

\[
d_{ij} = \int_{w<t} \tilde{C}^{11}\{F_{W_{ij}}(w), S_T(t)\} f_{W_{ij}}(w) f_T(t) dw dt
\]

is a normalizing constant, and \(\tilde{C}^{11}\) is the partial derivative of \(\tilde{C}\) with respect to both arguments. The copula \(\tilde{C}\) is a bivariate cdf on the unit square with uniform margins.

Model (4.9) uses copulas in a non-standard way since it features a cumulative distribution function and a survival function, which is referred to as the semi-survival copula by Chaieb-Lakhal et al. (2006). Note that \(F_{W_{ij}}(w)\) and \(1 - S_T(t)\) cannot be interpreted in terms of marginal distributions since they are associated with marginal behaviors of \(W_{ij}\) and \(T\) in the observable region (cf. Chaieb-Lakhal et al., 2006). Using joint survival copulas, (4.9) can be rewritten as

\[
R_{ij}(w, t) = \Pr\{W_{ij} \leq w, b - T_{ij} > t | b - T_{ij} > W_{ij}\} = \frac{1}{d_{ij}} \left[ S_T(t) - C\{S_{W_{ij}}(w), S_T(t)\} \right], \quad t \geq w, \quad (4.10)
\]

where \(C\) is a survival copula defined in Section 3.

Since \(I(W_{ij} \leq b - T_{ij})\) is a binomial variable with \(E[I(W_{ij} \leq b - T_{ij})] = d_{ij}\), it is convenient to parameterize it by

\[
d_{ij} = \exp(X_i^T \gamma)/[1 + \exp(X_i^T \gamma)],
\]

which emphasizes its dependence on covariates \(X_i\). Then we can apply the two-stage estimating procedures of Peng and Fine (2007) to our model (4.10) as follows.

To estimate \(\{\eta(t), \theta(t), \gamma\}\), for \(j = 1, \ldots, \tilde{m}_i\), where \(\tilde{m}_i\) is the number of observed pairs \((b - T_{ij}, W_{ij}), j = 1, \ldots, m_i\), proceeding along with the notations \(\eta(t)\) and \(\theta(t)\) in
Section 4.2, we consider

\[ B_{ij} = B_{ij} \{ \theta(t), \eta(t), \gamma, S_T(t), t \} \]

\[ = V_{ij} \tilde{D}_{ij} [I(Y_{ij} > t | b - T_{ij} > W_{ij}) - I(b - T_{ij} > t | b - T_{ij} > W_{ij}) \Phi_{ij}], \]

where \( \theta(t) = \theta(t, t) \), \( V_{ij} = V_{ij} \{ \theta(t), \eta(t), \gamma, t \} \) is a scalar weight function,

\[ \Phi_{ij} = \Phi_{ij} (t, w, \theta, \gamma) = S_T(t) - \frac{C(S_T(t), S_{W_{ij}}(w), \theta)}{S_T(t) [\exp(X_{ij}^T \gamma) / [1 + \exp(X_{ij}^T \gamma)]],} \]

and

\[ \tilde{D}_{ij} = \tilde{D}_{ij} \{ \theta(t), \eta(t), \gamma, S_T(t) \} = \frac{\partial \Phi_{ij} \{ \theta(t), \eta(t), \gamma, S_T(t) \}}{\partial \{ \theta(t), \eta(t), \gamma \}}. \]

Under the model assumptions, \( B_{ij} \) has zero mean conditional on \( X_3, I(b - T_{ij} > t) \) and \( b - T_{ij} > W_{ij} \) since \( E[I(Y_{ij} > t) | b - T_{ij} > t, b - T_{ij} > W_{ij}, X_3] = \Phi_{ij} \).

This is a nonlinear binary regression model for the effects of covariate \( X_2 \) on response \( I \{ \min(b - T_{ij}, W_{ij}) > t \} \) given \( W_{ij} > t \) and \( b - T_{ij} > W_{ij} \). Substituting an estimator \( \hat{S}_T(t) \) for \( S_T(t) \) and averaging over the \( m_i \) observations yield the estimating equation

\[ \sum_{i=1}^{n} \bar{U}_{ij}(\eta(t), \theta(t), \gamma, S_T(t), t) = \sum_{i=1}^{n} \sum_{j=1}^{\tilde{m}_i} B_{ij} (\eta(t), \theta(t), \gamma, S_T(t), t). \tag{4.11} \]

Estimation of \( S_T(t) \) based on the left-truncated data \( b - T_{ij} \) by \( W_{ij} \) can be found in Tsai et al. (1987). Although the data are correlated, computationally the conditional likelihood has the form of nonparametric likelihood for independent left-truncated data (Huang and Wang, 2004). Denote the data by \( \{ (b-t_{ij}, W_{ij}), j = 1, \ldots, \tilde{m}_i, i = 1, \ldots, n \} \) and let \( b - t_{(1)} < b - t_{(2)} < \cdots < b - t_{(\tilde{m}_i)} \) represent the distinct ordered values of \( \{ b - t_{ij}, j = 1, 2, \ldots, \tilde{m}_i, i = 1, \ldots, n \} \). Then the nonparametric maximum likelihood estimator of \( S_T(t) \) is given by a product-limit (cf. Wang, Jewell and Tsai, 1986):

\[ \hat{S}_T(t) = \prod_{l:b - t_{(l)} < t} \left( 1 - \frac{d_{(l)}}{R_{(l)}} \right), \tag{4.12} \]

where \( d_{(l)} \) is the number of \( b - t_{ij} \) equal to \( b - t_{(l)} \), and \( R_{(l)} \) is the total number of \( b - t_{ij} \) such that \( \{ b - t_{ij} \geq b - t_{(l)} \geq W_{ij} \} \).

Before closing this section, we would like to point out that, for either the IBNYR or the IBNER model, if the data include at least one claim reported before the endpoint \( b \) of observation, then the method proposed in this paper can be applied. Moreover, if the sample size is similar to that of Peng and Fine (2007, 2006), then our method has similar performance to theirs in the cases with censoring and truncation.
Remark 1. Because \((b - T_{ij}, W_{ij})\) are not always observable due to truncation, the interpretation of \(S_W(w)\) as the marginal survival function of \(W\) in (4.7) and (4.9) is controversial (Chaieb-Lakhal et al., 2008). Hence the standard two-steps procedure for copula model is not justified for this case. Conventional approach is to give an ad hoc estimate of the survival function for the terminating event \(b - T\), and then to estimate the copula parameter and the survival function of the non-terminating variable \(W\), such as the proportional hazards structure in Peng and Fine (2007) and the nonparametric assumption of the survival functions in Chaieb-Lakhal et al. (2006, 2008).

Remark 2. For semi-competing risks model, the estimators of \(\Lambda_0(t)\) are based on the empirical process of the observed event times \(\{t_{ij}, j = 1, 2, \ldots, m_i; i = 1, 2, \ldots, n\}\). For the case of semi-survival copula model, an estimator of \(\Lambda_0(t)\) can be attained similarly to (4.12) but with the observed \(\{t_{ij}, j = 1, 2, \ldots, \tilde{m}_i; i = 1, 2, \ldots, n\}\) (not \(\{b - t_{ij}\}\)).

Note that \(E[m_i|X_{i1}] = \exp(X_{i1}^T \beta) \Lambda_0(a)\) implies
\[
E[m_i \Lambda_0^{-1}(a)|X_{i1}] = \exp(X_{i1}^T \beta).
\]

Hence a class of estimating equations for \(\beta\) is defined by
\[
\sum_{i=1}^{n} \tilde{w}_i X_{i1}^T (m_i \Lambda_0^{-1}(a) - \exp(X_{i1}^T \beta)) = 0, \quad (4.13)
\]
where \(\tilde{w}_i\) is a weight function depending on \((X_{i1}, \beta, \Lambda_0)\). For Case I of censoring scheme, the estimate of \(\beta\) can be obtained by solving the estimating equation in (4.13) with \(\hat{\Lambda}_0(a)\) in place of \(\Lambda_0(a)\).

For the case of semi-survival copula model, since
\[
E \left[ \frac{\tilde{m}_i}{d_{ij}} \left| X_{i1} \right. \right] = E \left[ \frac{\tilde{m}_i}{\exp(\gamma^T X_{ij4})} \left| X_{i1} \right. \right] = \exp(X_{i1}^T \beta) \Lambda_0(a),
\]
a class of estimating equations for \(\beta\) is defined similarly to (4.13) as
\[
\sum_{i=1}^{n} \tilde{w}_i X_{i1}^T \left( \sum_{j=1}^{\tilde{m}_i} \frac{\Lambda_0^{-1}(a)[1 + \exp(X_{ij4}^T \gamma)]}{\exp(X_{ij4}^T \gamma)} - \exp(X_{i1}^T \beta) \right) = 0. \quad (4.14)
\]

The estimate of \(\beta\) can be obtained by solving the estimating equation in (4.14) with \(\hat{\Lambda}_0(a)\) and \(\hat{\gamma}\) in place of \(\Lambda_0(a)\) and \(\gamma\) respectively.
5. Asymptotic properties of estimators

If the pairs \((b - T_{ij}, W_{ij})\) are subject to censoring by an independent random variable \(C_{ij}\) and \(W_{ij}\) is right censored by \(b - T_{ij}\), the large sample properties of the estimators of the nonparametric component \(H_0(t)\) and the parametric component \(\alpha\) have been established in Peng and Fine (2007). Their results can be extended easily to the semi-competing risks model proposed in Section 4.2.

For semi-survival copula model of Section 4.3, we extend the results of Peng and Fine (2007) based on semi-competing risks model to the estimating equation defined in (4.11) based on left-truncated observations \((b - T_{ij}, W_{ij})\) if \(W_{ij} \leq b - T_{ij}\). Then the asymptotic properties presented in Peng and Fine (2007) still hold. To save space, we only state the asymptotic results of (4.11) below. The regularity conditions and the full proofs are relegated to Appendix.

**Theorem 1.** Under the model given by (2.3)-(2.4) and (4.5) and conditions \((C_1 - C_4)\) in Appendix, if \(\hat{S}_T(t)\) is a uniformly consistent estimator of the survival function \(S_T(t)\) of variable \(b - T_{ij}\) for \(t \in [l, u]\), then for \(n\) large enough, there exists a uniformly bounded solution \((\hat{\theta}(t), \hat{\eta}(t), \hat{\gamma})\) to equation \(\bar{U} = \sum_{i=1}^n \bar{U}_i \{\theta, \eta(t), \gamma, S_T(t), t\} = 0\) such that

\[
\sup_{t \in [l, u]} \left\| (\hat{\theta}(t), \hat{\eta}(t), \hat{\gamma})^\top - (\theta(t), \eta(t), \gamma)^\top \right\| \to 0 \quad \text{in probability.}
\]

**Theorem 2.** Assume that the conditions of Theorem 1 and \((C_5)\) in Appendix hold and there exist i.i.d. random functions \(\{\phi_i(t)\}_{i=1}^\infty\) such that

\[
\left\| \sup_{t \in [l, u]} \{\hat{S}_T(t) - S_T(t)\} - n^{-1/2} \sum_{i=1}^n \phi_i(t) \right\|_p \to 0
\]

and \(\{\phi_1(t), t \in [l, u]\} \) is Glivenko-Cantelli and Donsker. Then \(\sqrt{n}\{\hat{\theta} - \theta, \hat{\eta} - \eta, \hat{\gamma} - \gamma\}^\top\) converges weakly to a zero-mean Gaussian process with covariance function

\[
\Sigma(s, t) = \lim_{n \to \infty} \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^{\tilde{m}_i} E[G_{ij}(s)G_{ij}(t)^\top],
\]

where \(m = \sum_{i=1}^n \tilde{m}_i\),

\[
G_{ij}(t) = J(t)^{-1} [B_{ij} \{\theta(t), \eta(t), \gamma, S_T(t), t\} - H(t)\phi_i(t)], i = 1, 2, \ldots, n,
\]

and \(J(t)\) is the asymptotic limit of

\[
\hat{J}(t) = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^{\tilde{m}_i} V_{ij} \{\hat{\theta}(t), \hat{\eta}(t), \hat{\gamma}, t\} \left[\tilde{D}_{ij} \{\hat{\theta}(t), \hat{\eta}(t), \hat{\gamma}, \hat{S}_T(t)\}\right]^\otimes2 I(b - T_{ij} > t).
\]
Suppose there exist \( \hat{\phi}_i(t) \) such that \( \sup_{t \in [l, u]} \| \hat{\phi}_i(t) - \phi_i(t) \| \overset{P}{\to} 0 \) and let

\[
\hat{G}_{ij}(t) = \hat{J}^{-1}(t) \left[ B_{ij} \{ \hat{\theta}, \hat{\eta}, \hat{\gamma}, \hat{S}_T, t \} - \hat{H}(t) \hat{\phi}_i(t) \right], \quad \hat{\Sigma}(s, t) = (nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \hat{G}_{ij} \hat{G}_{ij}^T,
\]

where \( \hat{H}(t) \) is the asymptotic limit of

\[
\hat{H}_n(t) = (nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} V_{ij} \{ \hat{\theta}, \hat{\eta}, \hat{\gamma}, t \} D_{ij} \{ \hat{\theta}, \hat{\eta}, \hat{\gamma}, \hat{S}_T \} \hat{D}_{ij}^* \{ \hat{\theta}, \hat{\eta}, \hat{\gamma}, \hat{S}_T \} I(b - T_{ij} > t)
\]

with \( \hat{D}_{ij}^* \{ \theta, \eta, \gamma, S_T \} = \partial \Phi_{ij} \{ \theta, \eta, \gamma, S_T \} / \partial S_T \). Then \( \sup_{t \in [l, u]} \| \hat{\Sigma}(s, t) - \Sigma(s, t) \| \overset{P}{\to} 0 \).

6. Simulation results

In this section we report some simulation results to evaluate the finite-sample properties of the inference procedures of the model proposed in Section 4. Based on the margins given in (2.3) and (2.4), we can construct a new copula function with these given margins. Note that the margin defined in (2.3) can be approximated by proportional hazard structure with zero coefficients of covariates. Hence we can take similar margins in our simulation to those in Peng and Fine (2007).

In our simulations, the exposure interval is taken to be \((0, 2]\) and the reported events are in observation interval \((0, 2]\) (cf. Jewell, 1990, p.42), that is, \( a = b = 2 \). For each subject \( i \), we consider the case where covariate vectors \( X_{i1} \) and \( X_{i2} \) are time-independent, \( X_{i1} \) is uniformly distributed over interval \([0, 2]\), while \( X_{i2} = X_{i4} \) are independent normal variables with mean 1 and variance 0.5, constrained to \([0, 2]\).

For model (2.1) and (2.3), the baseline marginal rate function is \( \lambda_0(t) = t/2, t \in [0, 2] \) (hence \( \Lambda_0(2) = 1 \)), and the covariate coefficient is \( \beta = 1.5 \) for \( X_{i1} \). For model (2.4), the baseline distribution function \( F_W(w) = 1 - \exp(-\phi w) \) of the delays is defined by the baseline hazard function \( h_0(w) = \phi \) with \( \phi = 1 \), and the covariate coefficient is \( \alpha = 2 \) for \( X_{i2} \). Thus given \( X_{i2} = x_{i2} \), the delay for the event occurred at \( t_{ij} \) has a hazard rate function \( h_i(w) = h_0(w) \exp(x_{i2}^T \alpha) \). The recurrent event number \( m_i \) is generated from a non-stationary Poisson process with rate function \( \lambda_i(t) = \lambda_0(t) \exp(x_i^T \beta) \). Thus given \( (x_{i1}, z_i, m_i), m_i \geq 1 \), the event times \( \{ t_{i1}, t_{i2}, \ldots, t_{im_i} \} \) are the order statistics of i.i.d. random variables with density \( \pi(t) \) given in (2.3).

For the survival functions of \( b - T \) and \( W_{ij} \), we take \( S_T(t) = \Lambda_0(b - t) = (2 - t)^2/4, t \in (0, 2] \) and \( S_{W_{ij}}(w) = [1 - F_W(w)]^{\exp(x_{i2}^T \alpha)} = [\exp(-\phi w)]^{\exp(x_{i2}^T \alpha)} \) respectively, so
that \( \log \{ -\log S_{W_{ij}}(w) \} = \alpha^T X_{ij} + \log(\phi w) \). We generate \( e_{i,1} = \log \{ -\log S_T(b - T_{ij}) \} \) and \( e_{i,2} = \log \{ -\log S_{W_{ij}}(W_{ij}) \} = \alpha^T X_{ij} + \log(\phi W_{ij}) \), for subject \( i \), \( i = 1, \ldots, n \), with \( \Pr(e_{i,1} > t) = \Pr(e_{i,2} > t) = S(t) = \exp\{-\exp(t)\} \). The joint distribution of \((e_{i,1}, e_{i,2})\) is modeled by the gamma frailty copula
\[
\Pr(e_{i,1} > t_1, e_{i,2} > t_2) = [S(t_1)^{-\theta} + S(t_2)^{-\theta} - 1]^{-1/\theta}, \quad \theta > 0,
\]
which has the marginal distributions \( \Pr(e_{i,1} > t) = \Pr(e_{i,2} > t) = S(t) \). The marginal distributions of \( b - T \) and \( W_{ij} \) are then given by
\[
\Pr(b - T_{ij} > t) = \Pr(S_T(b - T_{ij}) < S_T(t)) = \Pr(e_{i,1} > \log \{ -\log S_T(t) \}) = S(\log \{ -\log S_T(t) \}) = \exp\{-\exp(\log \{ -\log S_T(t) \})\} = S_T(t)
\]
and similarly, \( \Pr(W_{ij} > w) = \Pr(e_{i,2} > \log \{ -\log S_{W_{ij}}(w) \}) = S_{W_{ij}}(w) \) as required. In the simulation, we take \( \theta(t) = \theta = 0.5 \) and \( \eta(t) = (\log(H_0(t)), \alpha)^T = (\log(\phi t), \alpha)^T, \ t > 0 \).

An ad hoc estimator for \( S_T(t) \) is its empirical distribution. For semi-survival copula model, since \( \Pr(b - T_{ij} > W_{ij}) = d_{ij} \), we can take \( \gamma = \logit(d_{ij})/x_{ij4} \).

Let \( F(t) = 1 - S(t) \) and define \( \bar{C}(\cdot, \cdot) \) by
\[
\bar{C}(u, v) = [(1 - u)^{-\theta} + (1 - v)^{-\theta} - 1]^{-1/\theta} + u + v - 1.
\]
Then the joint cdf of \((e_{i,1}, e_{i,2})\) can be obtained from (6.1)–(6.2) by
\[
\Pr(e_{i,1} \leq t_2, e_{i,2} \leq t_2) = \Pr(e_{i,1} > t_1, e_{i,2} > t_2) + \Pr(e_{i,1} \leq t_1) + \Pr(e_{i,2} \leq t_2) - 1
= [S(t_1)^{-\theta} + S(t_2)^{-\theta} - 1]^{-1/\theta} + F(t_1) + F(t_2) - 1 = \bar{C}(t_1, t_2).
\]
An algorithm to sample from (6.3) is given as follows: (1) generate a random variable \( u \) from the uniform distribution over \((0, 1) \) \((U(0, 1))\); (2) generate a random variable \( q \) from \( U(0, 1) \); and (3) calculate \( v = \bar{C}_u^{-1}(q|u) \), where \( \bar{C}_u(v|u) = \partial \bar{C}(u, v)/\partial u \). Then \((e_{i,1}, e_{i,2}) = (F^{-1}(u), F^{-1}(v))\) is the desired sample with the joint distribution in (6.3).

The simulation results are reported in Table 1 and Table 2 below, where \( m(\cdot) \) and \( d(\cdot) \) denote the means and standard deviations of the estimated parameters of \( \alpha, \gamma \) and \( \theta(t), H_0(t) \) at points \( t_l \in [0, 2] \) with \( t_l = l(2 - 0.1)/8, \ l = 1, 2, \ldots, 8 \). The weight function \( V_{ij} \) is set to 1 in all simulations.

From Tables 1 and 2, we can see that the estimated parameters perform well. For semi-competing risk model (4.3), an estimator of \( \beta \) is also given by 0.1503 with the estimated variance 0.0243 under the sample size \( n = 200 \) based on the estimating equation (4.8); for semi-survival copula model, an estimator of \( \beta \) is 0.1658 with the estimated variance 0.0365 under the sample size \( n = 200 \) based on the estimating equation (4.9).
**Table 1** Summary of the simulation studies for model (4.3)

<table>
<thead>
<tr>
<th>n</th>
<th>t</th>
<th>m(α)</th>
<th>d(α)</th>
<th>m(H₀(t))</th>
<th>d(H₀(t))</th>
<th>m(θ)</th>
<th>d(θ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>t₁</td>
<td>2.076</td>
<td>0.447</td>
<td>0.319</td>
<td>0.142</td>
<td>0.425</td>
<td>0.365</td>
</tr>
<tr>
<td></td>
<td>t₂</td>
<td>1.861</td>
<td>0.708</td>
<td>0.421</td>
<td>0.411</td>
<td>0.463</td>
<td>0.326</td>
</tr>
<tr>
<td></td>
<td>t₃</td>
<td>2.231</td>
<td>0.379</td>
<td>0.537</td>
<td>0.971</td>
<td>0.534</td>
<td>0.633</td>
</tr>
<tr>
<td></td>
<td>t₄</td>
<td>1.938</td>
<td>0.303</td>
<td>0.940</td>
<td>0.515</td>
<td>0.578</td>
<td>0.850</td>
</tr>
<tr>
<td></td>
<td>t₅</td>
<td>1.949</td>
<td>0.347</td>
<td>1.256</td>
<td>0.427</td>
<td>0.433</td>
<td>0.173</td>
</tr>
<tr>
<td></td>
<td>t₆</td>
<td>2.423</td>
<td>0.287</td>
<td>1.522</td>
<td>0.920</td>
<td>0.437</td>
<td>0.599</td>
</tr>
<tr>
<td></td>
<td>t₇</td>
<td>2.169</td>
<td>0.618</td>
<td>1.693</td>
<td>0.798</td>
<td>0.343</td>
<td>0.298</td>
</tr>
<tr>
<td></td>
<td>t₈</td>
<td>1.930</td>
<td>0.908</td>
<td>1.896</td>
<td>0.896</td>
<td>0.484</td>
<td>0.148</td>
</tr>
<tr>
<td>200</td>
<td>t₁</td>
<td>1.982</td>
<td>0.268</td>
<td>0.336</td>
<td>0.268</td>
<td>0.465</td>
<td>0.343</td>
</tr>
<tr>
<td></td>
<td>t₂</td>
<td>1.957</td>
<td>0.428</td>
<td>0.429</td>
<td>0.403</td>
<td>0.524</td>
<td>0.303</td>
</tr>
<tr>
<td></td>
<td>t₃</td>
<td>1.987</td>
<td>0.630</td>
<td>0.698</td>
<td>0.258</td>
<td>0.556</td>
<td>0.223</td>
</tr>
<tr>
<td></td>
<td>t₄</td>
<td>1.970</td>
<td>0.548</td>
<td>0.961</td>
<td>0.324</td>
<td>0.529</td>
<td>0.222</td>
</tr>
<tr>
<td></td>
<td>t₅</td>
<td>1.976</td>
<td>0.259</td>
<td>1.210</td>
<td>0.319</td>
<td>0.420</td>
<td>0.280</td>
</tr>
<tr>
<td></td>
<td>t₆</td>
<td>1.955</td>
<td>0.296</td>
<td>1.432</td>
<td>0.472</td>
<td>0.346</td>
<td>0.173</td>
</tr>
<tr>
<td></td>
<td>t₇</td>
<td>2.053</td>
<td>0.507</td>
<td>1.677</td>
<td>0.401</td>
<td>0.385</td>
<td>0.228</td>
</tr>
<tr>
<td></td>
<td>t₈</td>
<td>2.041</td>
<td>0.601</td>
<td>1.956</td>
<td>0.298</td>
<td>0.439</td>
<td>0.198</td>
</tr>
</tbody>
</table>

**Table 2** Summary of the simulation studies for model (4.6)

<table>
<thead>
<tr>
<th>n</th>
<th>t</th>
<th>m(α)</th>
<th>d(α)</th>
<th>m(H₀(t))</th>
<th>d(H₀(t))</th>
<th>m(θ)</th>
<th>d(θ)</th>
<th>m(γ)</th>
<th>d(γ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>t₁</td>
<td>1.947</td>
<td>0.116</td>
<td>0.326</td>
<td>0.157</td>
<td>0.618</td>
<td>0.307</td>
<td>−0.213</td>
<td>0.528</td>
</tr>
<tr>
<td></td>
<td>t₂</td>
<td>1.911</td>
<td>0.517</td>
<td>0.490</td>
<td>0.416</td>
<td>0.419</td>
<td>0.626</td>
<td>−0.245</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td>t₃</td>
<td>1.533</td>
<td>0.812</td>
<td>0.561</td>
<td>0.812</td>
<td>0.609</td>
<td>0.199</td>
<td>−0.255</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td>t₄</td>
<td>1.942</td>
<td>0.186</td>
<td>0.988</td>
<td>0.581</td>
<td>0.578</td>
<td>0.592</td>
<td>−0.167</td>
<td>0.459</td>
</tr>
<tr>
<td></td>
<td>t₅</td>
<td>1.626</td>
<td>0.595</td>
<td>1.267</td>
<td>0.595</td>
<td>0.471</td>
<td>0.295</td>
<td>−0.216</td>
<td>0.560</td>
</tr>
<tr>
<td></td>
<td>t₆</td>
<td>1.856</td>
<td>0.187</td>
<td>1.678</td>
<td>0.871</td>
<td>0.587</td>
<td>0.263</td>
<td>−0.186</td>
<td>0.305</td>
</tr>
<tr>
<td></td>
<td>t₇</td>
<td>1.788</td>
<td>0.408</td>
<td>1.722</td>
<td>0.798</td>
<td>0.286</td>
<td>0.592</td>
<td>−0.177</td>
<td>0.382</td>
</tr>
<tr>
<td></td>
<td>t₈</td>
<td>2.044</td>
<td>0.102</td>
<td>1.951</td>
<td>0.102</td>
<td>0.493</td>
<td>0.256</td>
<td>−0.213</td>
<td>0.380</td>
</tr>
<tr>
<td>200</td>
<td>t₁</td>
<td>2.039</td>
<td>0.361</td>
<td>0.342</td>
<td>0.287</td>
<td>0.471</td>
<td>0.311</td>
<td>−0.240</td>
<td>0.597</td>
</tr>
<tr>
<td></td>
<td>t₂</td>
<td>1.945</td>
<td>0.655</td>
<td>0.447</td>
<td>0.413</td>
<td>0.450</td>
<td>0.258</td>
<td>−0.235</td>
<td>0.407</td>
</tr>
<tr>
<td></td>
<td>t₃</td>
<td>1.968</td>
<td>0.630</td>
<td>0.701</td>
<td>0.292</td>
<td>0.388</td>
<td>0.250</td>
<td>−0.256</td>
<td>0.427</td>
</tr>
<tr>
<td></td>
<td>t₄</td>
<td>1.970</td>
<td>0.494</td>
<td>0.965</td>
<td>0.311</td>
<td>0.297</td>
<td>0.357</td>
<td>−0.234</td>
<td>0.238</td>
</tr>
<tr>
<td></td>
<td>t₅</td>
<td>1.932</td>
<td>0.193</td>
<td>1.243</td>
<td>0.319</td>
<td>0.598</td>
<td>0.291</td>
<td>−0.229</td>
<td>0.459</td>
</tr>
<tr>
<td></td>
<td>t₆</td>
<td>1.971</td>
<td>0.193</td>
<td>1.441</td>
<td>0.493</td>
<td>0.356</td>
<td>0.162</td>
<td>−0.224</td>
<td>0.399</td>
</tr>
<tr>
<td></td>
<td>t₇</td>
<td>1.891</td>
<td>0.537</td>
<td>1.683</td>
<td>0.456</td>
<td>0.256</td>
<td>0.316</td>
<td>−0.272</td>
<td>0.316</td>
</tr>
<tr>
<td></td>
<td>t₈</td>
<td>1.848</td>
<td>0.542</td>
<td>1.976</td>
<td>0.287</td>
<td>0.404</td>
<td>0.190</td>
<td>−0.262</td>
<td>0.296</td>
</tr>
</tbody>
</table>
7. Concluding remarks

In this paper we review existing individual claim loss reserving models, and then propose semiparametric models for both the arrival process of the claims and the hazard function of the delay variables, which can accommodate covariates. The semi-competing risk model and the semi-survival copula model proposed in this paper are more flexible to predict the claim loss reserving. A nonstandard two-step estimating procedure is proposed to draw statistical inference, which is similar to pseudo likelihood estimation, and the large-sample properties of the estimators are provided as well. The simulation study indicates that the proposed procedure can produce efficient estimates and improve predictions for the claim loss reserving.

Goodness-of-fit tests are important for assessing the adequacy of the proposed model and related methods, which can be found in the literature such as Peng and Fine (2007). The kernel estimation and the sieve likelihood are proposed by Gijbels and Mielniczuk (1990), Bouezmarni and Rombouts (2009) and Chen and Huang (2007) under a semi-parametric copula with nonparametric margins. The local likelihood method is another nonparametric estimating procedure which has been illustrated to be superior to other nonparametric methods. Thus under the copula model, the local likelihood for margins with nonstandard two-step or simultaneous estimation approaches is of considerable interest and points to a future research direction.

Cure models (cf. Maller and Zhou, 1996) have been extensively discussed in survival analysis, which are also investigated in the actuarial literature in the form of zero-inflated model (cf. Boucher et al., 2007, Boucher and Denuit, 2007, 2008) for the arrival process of the claim counts, which implies a positive fraction of the subjects who never experience the event of interest. Furthermore, the claim occurred at time $t_{ij}$ for subject $i$ might be reported at once, so that the delay is zero. Thus the cure model may be applied to tackle the IBNR problems if there exists a positive fraction of claims which are reported immediately. These are important and challenging issues of interest for further studies.

Acknowledgment: We are grateful to an associate editor and referees for their valuable comments and suggestions that helped improve this paper substantially.
References


Appendix:

To prove Theorems 1–3, we need the following notations and regularity conditions. Define

\[
K_{ij}(\tau, \kappa, \varsigma, \rho, \theta(t), \eta(t), S_T(t), \gamma, t) = V_{ij}\{\tau(t), \kappa(t), \rho, t\} \tilde{D}_{ij}\{\tau(t), \kappa(t), \varsigma(t), \rho\} \times [I(Y_{ij} > t|b - T_{ij} > \tilde{W}_{ij}) - I(b - T_{ij} > \tilde{W}_{ij})\Phi_{ij}\{\theta(t), \eta(t), \gamma, S_T(t)\}],
\]

\[
L_{ij}(\tau, \kappa, \varsigma, \rho, \theta(t), \eta(t), S_T(t), \gamma, t) = V_{ij}\{\tau(t), \kappa(t), \rho, t\} \tilde{D}_{ij}\{\tau(t), \kappa(t), \varsigma(t), \rho\} \\
\quad \times \tilde{D}_{ij}^\top\{\theta(t), \eta(t), \gamma, S_T(t)\}I(b - T_{ij} > t),
\]

\[
Q_{ij}(\tau, \kappa, \varsigma, \rho, \theta(t), \eta(t), S_T(t), \gamma) = V_{ij}\{\tau(t), \kappa(t), \rho, t\} \tilde{D}_{ij}\{\tau(t), \kappa(t), \varsigma(t), \rho\} \\
\quad \times \tilde{D}_{ij}^*\{\theta(t), \eta(t), \gamma, S_T(t)\}I(b - T_{ij} > t),
\]

where

\[
L_{ij}(\tau, \kappa, \varsigma, \rho, \theta(t), \eta(t), S_T(t), \gamma) = E\{L_{ij}(\tau, \kappa, \varsigma, \rho, \theta(t), \eta(t), S_T(t), \gamma)\},
\]

\[
L(\tau, \kappa, \varsigma, \rho, \theta(t), \eta(t), S_T(t)) = (nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} L_{ij},
\]

and

\[
\tilde{D}_{ij}^*\{\theta(t), \eta(t), \gamma, S_T(t)\} = \frac{\partial\Phi_{ij}\{\theta(t), \eta(t), \gamma, S_T(t)\}}{\partial S_T(t)}.
\]

Regularity conditions:

\( (C_1) \) \sup_{t \in [l, u]} |\theta(t)| < \infty, \sup_{t \in [l, u]} \|\eta(t)\| < \infty, and parameter \( \gamma \) and covariates \( X_{i1}, X_{i2}, X_{i4} \) are bounded with probability 1.

\( (C_2) \) \( \Phi_{ij}(s, t, u, v) \) and all components of \( \partial\Phi_{ij}(s, t, u, v) / \partial(s, t, u, v) \) are Lipschitz continuous.

\( (C_3) \) \inf_{t \in [l, u]} \text{mineig}\{J(t)\} > 0, where \text{mineig} is the minimum eigenvalue of a matrix.

\( (C_4) \) Random weight functions \( V_{ij} \) (\( i = 1, 2, \ldots, n; j = 1, 2, \ldots, \tilde{m}_i \)) are independent of \( W_{ij} \) and \( b - T_{ij} \). For all bounded \( A \subset \mathcal{R} \), \( B \subset \mathcal{R}^{p+1} \) and \( D \subset \mathcal{R} \), the class of random functions \( \{V_{ij}(a, b, c, t), a \in A, b \in B, c \in D, t \in [l, u]\} \) is bounded below and above by positive constants and is a Donsker class (Van der Vaart and Wellner, 1996). In particular, if \( V_{ij}(a, b, c, t) \) is a nonrandom function and is Lipschitz continuous, then the class \( \{V_{ij}(a, b, c, t), a \in A, b \in B, c \in D, t \in [l, u]\} \) is Donsker.

\( (C_5) \) All components of \( \partial^2\Phi_{ij}(s, t, u, v) / \partial(s, t, u, v) \partial(s, t, u, v) \) are Lipschitz continuous. The class of random functions \( \{\partial V_{ij}(a, b, c, t) / \partial(a, b, c, t), a \in A, b \in B, c \in D, t \in [l, u]\} \) is bounded and is a Donsker class for all bounded \( A \subset \mathcal{R} \), \( B \subset \mathcal{R}^{p+1} \) and \( D \subset \mathcal{R} \).
Proof of Theorem 1: Write $K_{ij} = K_{ij}(\tau, \kappa, \zeta, \rho, \theta(t))$ and define

$$Q = \{K_{ij} : \tau, \theta \in \ell_c^\infty([l, u]), \rho, \gamma \in [d_1, d_2], \kappa, \eta \in \ell_c^\infty([l, u])\}^{p+1}\).$$

The class $Q$ is Glivenko-Cantelli and Donsker since $\{(t, \infty) : t \in [l, u]\}$ is a Donsker class; $\eta(t)\top X_{1,3}$ and $\eta(t) \in \ell_c^\infty([l, u])\}$, $t \in [l, u]$, are equivalent to the Donsker class $\{b_l X_{1,3}, b \in [-c, c]\}^{p+1}$; and $\rho$ and $\gamma$ are bounded in $[d_1, d_2]$; the set of all $S_T(t)$ is Donsker as $S_T(t)$ is monotone (cf. example 2.6.21 of Van der Vaart and Wellner 1996); Donsker property is preserved under Lipschitz transformation, sum and product operations; and every Donsker class is a Glivenko-Cantelli class in probability.

For any $\tilde{\theta} \in \ell_c^\infty([l, u]), \tilde{\eta}(t) \ell_c^\infty([l, u])\}$, $\tilde{\gamma} \in [d_1, d_2]$, one can easily show that

$$\frac{1}{nm} \tilde{U}\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T, t\} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^{m_i} (K_{ij}(\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T, \theta, \eta, \gamma, S_T, t) - V_{ij}\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, t\}
\cdot \tilde{D}_{ij}\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T\}I(b - T_{ij} > t)[\Phi_{ij}\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T\} - \Phi_{ij}\{\theta, \eta, \gamma, \tilde{S}_T\}],$$

since $\Phi_{ij}\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T\} - \Phi_{ij}\{\theta, \eta, \gamma, \tilde{S}_T\}$ is equivalent to

$$\tilde{D}_{ij}\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T\}^\top (\theta - \eta, \eta, \gamma - \gamma) + \tilde{D}_{ij}\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T\}^\top \{\tilde{S}_T - S_T\},$$

where $\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T\}$ is on the line segment between $\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T\}$ and $\{\theta, \eta, \gamma, \tilde{S}_T\}$. Hence

$$\frac{1}{nm} \tilde{U}\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T, t\} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^{m_i} (K_{ij}(\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T, \theta, \eta, \gamma, S_T, t) - V_{ij}\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, t\}
\cdot I(b - T_{ij} > t)\tilde{D}_{ij}\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T\}\tilde{D}_{ij}\{\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T\}^\top (\theta - \eta, \eta, \gamma - \gamma) + \varepsilon_n(t) + \varepsilon_n(t).$$

Since $\tilde{S}_T(t)$ is a uniformly consistent estimator, $\sup_{t \in [l, u]} \|\varepsilon_n(t)\| \overset{p}{\to} 0$ from conditions $(C_1)$, $(C_2)$ and $(C_4)$.

Since $Q$ is Glivenko-Cantelli and $E[K_{ij}(\tilde{\theta}, \tilde{\eta}, \gamma, \tilde{S}_T, \theta, \eta, \gamma, \tilde{S}_T, t)] = 0$, we have

$$\left\| \sup_{t \in [l, u]} \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^{m_i} K_{ij} (\tilde{\theta}, \tilde{\eta}, \tilde{\gamma}, \tilde{S}_T, \theta, \eta, \gamma, S_T, t) \right\| \overset{p}{\to} 0.$$
Proof of Theorems 2 and 3: Simple algebra shows that

\[
0 = (nm)^{-1/2} \tilde{U} \{ \hat{\theta}, \hat{\eta}, \hat{\gamma}, \hat{S}_T, t \} \\
= (nm)^{-1/2} \tilde{U} \{ \theta, \eta, \gamma, S_T, t \} + (nm)^{-1/2} \left[ \tilde{U} \{ \hat{\theta}, \hat{\eta}, \hat{\gamma}, \hat{S}_T, t \} - \tilde{U} \{ \theta, \eta, \gamma, S_T, t \} \right] \\
= \tilde{U}_0(t) + \tilde{U}_1(t) - \tilde{U}_2(t),
\]

where \( \tilde{U}_0(t) = (nm)^{-1/2} \tilde{U} \{ \theta, \eta, \gamma, S_T, t \} \), \( \tilde{U}_1(t) = (nm)^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{\tilde{m}_i} [K_{ij}(\hat{\theta}, \hat{\eta}, \hat{\gamma}; \theta(t), \eta(t), S_T(t), \gamma; \tau(t), \eta(t), S_T(t), \gamma), \tau(t), \eta(t), S_T(t), \gamma)] \), and

\[
\tilde{U}_2(t) = (nm)^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{\tilde{m}_i} V_{ij} \{ \hat{\theta}, \hat{\eta}, \hat{\gamma}, \hat{S}_T, \gamma \} I(b - T_{ij} > t) \\
\cdot [\Phi_{ij}(\hat{\theta}(t), \hat{\eta}(t), \hat{\gamma}, \hat{S}_T) - \Phi_{ij}(\theta(t), \eta(t), \gamma, S_T(t))].
\]

Let \( \tilde{D}_{ij} \{ \theta, \eta, \gamma, S_T \} = \partial \{ V_{ij} \{ \theta, \eta, \gamma, S_T \} \tilde{D}_{ij} \{ \theta, \eta, \gamma, S_T \} \} / \partial \{ \theta, \eta, \gamma, S_T \}^\top \). From a Taylor expansion of \( V_{ij} \{ \theta_0, \eta_0, \gamma_0, t \} \tilde{D}_{ij} \{ \theta_0, \eta_0, \gamma_0, S_T \} \) at \( \{ \theta, \eta, \gamma, S_T \} \),

\[
\tilde{U}_1(t) = (nm)^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{\tilde{m}_i} \tilde{D}_{ij} \{ \theta, \eta, \gamma, S_T \} [I(Y_{ij} > t) - I(b - T_{ij} > t) \Phi_{ij}(\theta, \eta, \gamma, S_T)] \\
\cdot \sqrt{nm} \{ \hat{\theta} - \theta, \hat{\eta} - \eta, \hat{\gamma} - \gamma, \hat{S}_T - S_T \}^\top + v_n(t).
\]

Uniform consistency of \( \{ \hat{\theta}, \hat{\eta}, \hat{\gamma}, \hat{S}_T \} \) and condition (C5) imply \( \sup_{t \in [l, u]} \|v_n(t)\| \xrightarrow{p} 0 \) and \( \tilde{D}_{ij} \{ \theta, \eta, \gamma, S_T \} [I(Y_{ij} > t) - I(b - T_{ij} > t) \Phi_{ij}(\theta, \eta, \gamma, S_T)] \) is Glivenko-Cantelli over \( t \in [l, u] \). Thus \( \tilde{U}_1(t) = v_n(t) \sqrt{nm} \{ \hat{\theta} - \theta, \hat{\eta} - \eta, \hat{\gamma} - \gamma, \hat{S}_T - S_T \}^\top \) with \( \sup_{t \in [l, u]} \|v_n(t)\| \xrightarrow{p} 0 \).

Next by applying Taylor expansion and uniform law of large number,

\[
\tilde{U}_2(t) = \{ J(t) + v_n^*(t) \} \sqrt{nm} \{ \hat{\theta} - \theta, \hat{\eta} - \eta, \hat{\gamma} - \gamma \}^\top + \{ H(t) + v_n^{**}(t) \} \sqrt{nm} \{ \hat{S}_T - S_T \},
\]

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

\[
\sqrt{nm} \{ \hat{\theta} - \theta, \hat{\eta} - \eta, \hat{\gamma} - \gamma \}^\top = \sqrt{nm} J^{-1}(t) [\tilde{U} \{ \theta, \eta, \gamma, t \} - H(t) \{ \hat{S}_T - S_T \}] + \tau_n(t),
\]

where \( \sup_{t \in [l, u]} \|\tau_n(t)\| \xrightarrow{p} 0 \). Weak convergence follows since \( \{ B_{ij} \theta, \eta, \gamma, S_T \} \) is a subclass of \( Q \), and \( \{ \phi_i(t) : t \in [l, u] \} \) is a Donsker class.

Uniform consistency of \( \{ \hat{J}(t), \hat{H}(t) \} \) for \( \{ J(t), H(t) \} \) follows from Glivenko-Cantelli property of \( Q^* \) and \( Q^{**} \) and uniform consistency of \( \theta, \eta, \gamma, S_T \). Since \( \{ B_{ij} \theta, eta, \gamma, S_T \} \) and \( \{ \phi_i(t) : t \in [l, u] \} \) are Donsker classes, and \( J^{-1}(t) \) and \( H(t) \) are bounded for \( t \in [l, u] \), \( \{ G_{ij} : t \in [l, u] \} \) is also Glivenko-Cantelli. By Slutsky’s theorem and uniform law of large numbers, \( \hat{S}(s, t) \) converges to \( \Sigma(s, t) \) uniformly.