Competing Risks Regression with Dependent Multiple Spells

Monte Carlo Evidence and an Application to Maternity Leave

Lipowski, Cäcilia; Lo, Simon M. S. ; Shi, Shuolin ; Wilke, Ralf A.

Document Version
Accepted author manuscript

Published in:
Japanese Journal of Statistics and Data Science

DOI:
10.1007/s42081-021-00110-6

Publication date:
2021

Competing Risks Regression with Dependent Multiple Spells: Monte Carlo Evidence and an Application to Maternity Leave

Cäcilia Lipowski*
Simon M.S. Lo†
Shuolin Shi‡
Ralf A. Wilke§

January 27, 2021

Abstract

Copulas are a convenient tool for modelling dependencies in competing risks models with multiple spells. This paper introduces several practical extensions to the nested copula model and focuses on the choice of the hazard model and copula. A simulation study looks at the relevance of the assumed parametric or semiparametric model for hazard functions, copula and whether a full or partial maximum likelihood approach is chosen. The results show that the researcher must be careful which hazard is being specified as similar functional form assumptions for the subdistribution and cause-specific hazard will lead to differences in estimated cumulative incidences. Model selection tests for the choice of the hazard model and copula are found to provide some guidance for setting up the model. The nice practical properties and flexibility of the copula model are demonstrated with an application to a large set of maternity leave periods of mothers after they have given birth to up to their third child.
Keywords: Copula, Competing Risks, Repeated Occurrences, Maternity Leave

*ZEW Mannheim, E–mail: Caecilia.Lipowski@zew.de
†City University Hong-Kong, Department of Economics, E–mail: losimonms@yahoo.com.hk
‡Copenhagen Business School, Department of Economics, E–mail: ssh.eco@cbs.dk
§Copenhagen Business School and ZEW, Department of Economics, E–mail: rw.eco@cbs.dk

1 Introduction

Classical analyses of failure time, duration or time to event data assume a random sample of observations. Non-observation of the failure time is often assumed to be the result of independent censoring, such as end of observation period. These assumptions may be incorrect when the same unit generates more than one observed duration and if there are multiple causes of failure or competing risks. Risk and spell dependencies can be conveniently modelled with the help of a copula structure. Copula models for risk dependencies have been introduced by Carrièrre (1994), Zheng and Klein (1995) and Rivest and Wells (2001). Various extensions to models with truncation, covariates and multiple risks have been suggested by de Uña-Álvarez and Veraverbeke (2017), Braekers and Veraverbeke (2005), Lo and Wilke (2010, 2014) and Ha et al. (2019) among others. For an extensive treatment of copula modelling with dependent risks see Emura et al. (2019). A nested copula model that allows for additional dependencies between multiple spells or repeated occurrences has been suggested by Lo, Mammen and Wilke (2020). This paper bases on the latter model and introduces several practically relevant extensions such as additional model components that vary across multiple spells and flexible specifications of the hazard model. The piecewise constant hazard model (Colvert and Boardman, 1976; Lancaster, 1990) is a popular, because flexible model for the marginal hazard in applications. By dividing the time axis into fixed intervals, it is capable of approximating unknown hazard functions by a sequence of constants when intervals are short enough. Despite its flexibility it is a parametric model and is conveniently estimated by maximum likelihood. Although this modelling approach has gained some popularity for the marginal hazard function, the equivalent for the cause-specific hazard (Lawless, 2003) has not been adopted frequently (prominent exceptions are Craiu and Duchesne, 2004; Craiu and Lee, 2005; Kyyrä, 2009). The standard route in applied research is to specify the related subdistribution hazard by assuming a semiparametric proportional hazard model (Fine and Gray, 1999). This model is implemented in the main statistical packages and has a Google citation count of more than 8,500. There are therefore three different hazard functions that can be modelled in an application: the marginal hazard, the cause-specific hazard and the subdistribution hazard. The link between them is elaborated in detail in Emura et al. (2020). While the marginal hazard can only be identified under strong and non-testable additional restrictions on the model, the cause-specific and subdistribution hazard are identified and can be more easily estimated. As pointed out by Bakoyannis and Touloumi (2012), the latter two can be used to estimate cumulative incidence curves but the interpretation of the subdistribution hazard is less natural than the cause-specific hazard. Therefore, for an applied analysis it is more intuitive to work with the latter.

This paper extensively studies practical properties of the copula model, including finite sample performance and applicability to large data. While previous numerical analyses had an illustrative nature and were restricted to simple parametric examples with time constant covariates, little is known how the nested copula model performs in wider settings with more flexibility in the model. This includes the setting when model components are allowed to change over multiple spells, the choice of the hazard function or copula. It is shown that specifying similar proportional hazards models for the cause-specific and subdistribution hazard can give substantially different estimates for cumulative incidences. It therefore confirms the findings of Beyersmann et al. (2009) in a more general setting with multiple spells. As an addition to the literature (e.g. Beyersmann et al., 2009; Bakoyannis and Touloumi, 2012; Emura et al., 2020), we show that modelling the cause-specific hazard is compatible with a richer pattern of covariate effects on cumulative incidences, in particular that the effect can have opposite directions at different durations. We present direct comparisons of different hazard models and comparative numerical analyses by simulating a competing risks model with multiple spells and partly spell varying regressors. The analysis also includes a comparison of the performance of a partial maximum likelihood approach that omits spell dependencies and a full maximum likelihood approach. Additionally, the numerical behaviour of model selection tests for non-nested models (Vuong, 1989) is analysed, which include the choice of the functional form of the conditional hazard or the copula. Both are important model ingredients which are typically unknown in applications and, therefore, a data based guidance is preferable. It is also shown with large scale administrative maternity leave data that the piecewise constant nested copula model gives insightful results in applications that require a flexible shape of the hazard function due to the existence of mass points. In comparison to the semiparametric subdistribution hazard model it has the advantage that the baseline cause-specific hazard has a natural interpretation and can be visually inspected.

In our application we study the out of work duration of females who have given birth to a child. Fertility rates below reproduction level are a severe threat to the future of a range of highly developed countries in Eastern Asia, Europe and North America. Particular prominent examples in Asia are South Korea (Birthrate: 1.0), Hong Kong (1.1), Taiwan (1.2) and Japan (1.4). The social sciences literature (Waldfogel, 1997, Budig and England, 2001, Gangl and Ziefle, 2009) blames child bearing and rearing to be a main reason for lower earnings of females compared to males. In order to substantially increase fertility carefully designed policies are required. These can only be developed with a profound understanding of the factors why and when mothers return to the labour market after giving birth. Empirical analyses of maternity leave duration are still pretty scarce (examples are Fitzenberger et al., 2016; Arntz et al., 2017; Rodrigues and Vergnat, 2019) and restricted to the first child. There is currently no study looking into multiple spells due to giving birth to more than one child, although this multiple spell setting is the relevant scenario in order to avoid that a population becomes extinct. Hence, our empirical analysis of maternity leave focuses on up to the first three children by applying a range of flexible specifications of the copula duration model. Our data are linked social security records from Germany and contain in total maternity leave data from 34,380 mothers. Our results suggest that covariate effects of the determinants of the length of maternity leave strongly differ across competing risks and multiple spells. By specifying a flexible model for the hazards, we are able to fit well the data with mass points at certain durations which are due to the institutional framework.

Sample code for the models that are used in the numerical analysis in this paper can be downloaded from: `https://github.com/ralfawilke/copulamultiplespells`

The structure of the paper is as follows. The second section introduces the model. The third section presents results from simulation studies before the application results are presented in the fourth section. The last section summarises the main findings and gives ideas for future work.

2 Model

We consider a copula model for multiple spells and multiple risks duration data. To simplify the notation, we discuss a two-risks-K-spells model. Generalisation to more than two risks is straightforward by using the risk pooling method (Lo and Wilke, 2010). In our numerical analysis we restrict the model to $K = 3$ to reduce complexity for the reasons outlined in Lo et al. (2020).

Let T_{jk} be the latent duration for risk j in the k-th spell of a unit or agent, with $j = 1, 2$ and $k = 1, \ldots, K$. In the competing risks setting, latent durations are not observable, but the minimum of them, $T_k = \min\{T_{1k}, T_{2k}\}$, and the risk indicator $\Delta_k = \operatorname{argmin}\{j : T_{jk}\}$. The distribution of the observed tuples (T_k, Δ_k) for spell $k = 1, \ldots, K$ is described by the cumulative incidence curves (CIC)

$$Q_j(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) = \Pr(T_k \leq t, \Delta_k = j | \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}), \quad j = 1, 2, \tag{1}$$

where $\boldsymbol{x}_k$ is a $L_k \times 1$ vector of spell specific observed covariates. In our model the covariates differ across spells. This could be due to variation of the value but also due to different variables. $\boldsymbol{\beta}_{jk}$ are $M \times 1$ vectors of unknown parameters. These include shape parameters of Q_j and the parameters on the covariates. Therefore, this model allows regressors and their parameters to vary across spells, but covariates are constant within spells. Let $\boldsymbol{T} = (T_1, \ldots, T_K)'$, $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_K)'$, $\boldsymbol{x} = (\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_K)'$, $\boldsymbol{\beta}_k = (\boldsymbol{\beta}'_{1k}, \boldsymbol{\beta}'_{2k})'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_K)'$.

Let $S(t; \boldsymbol{x}_k, \boldsymbol{\beta}_k) = \Pr(T_k > t | \boldsymbol{x}_k, \boldsymbol{\beta}_k)$ be the overall survival function. By definition,

$$S(t; \boldsymbol{x}_k, \boldsymbol{\beta}_k) = 1 - Q_1(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{1k}) - Q_2(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{2k}). \tag{2}$$

Identifiability of $S(t; \boldsymbol{x}_k, \boldsymbol{\beta}_k)$ follows from identifiability of Q_j for $j = 1, 2$. In order to identify additional model components, one needs, in addition to Q_j for $j = 1, 2$, an assumed dependence structure between competing risks (Zheng and Klein, 1995; Rivest and Wells, 2001). This is true for both single spell and multiple spells models. Many empirical works with single spell models restrict their focus on Q_j to avoid making intestable assumptions about the dependence structure (Bakoyannis and Touloumi, 2012). In a multiple spells setting, the additional dependencies stemming from repeated observations also play a role for the joint distribution of durations, which are required to set up the likelihood. There are therefore two dependencies in our model: risk dependence and multiple spells dependence, which can both be characterised by a copula. To model this, we adopt a nested copula structure with the competing risks dependence as the daughter copula C_D and the multiple spells dependence as the mother copula C_M. Let $S_{jk}(t_{jk}; \boldsymbol{x}_k) = Pr(T_{jk} > t_{jk}|\boldsymbol{x}_k, \boldsymbol{\beta}_k)$ be the marginal survival for risk $j = 1, 2$ in spell $k = 1, \ldots, K$. For the usual reasons (i.e. copula-graphic estimator), S_{jk} can be identified under additional restrictions as a function of the CICs for the two risks. We therefore write it as a function of the parameters β_k. According to Sklar's theorem (Sklar, 1959), the joint distribution of two competing latent durations for the k^{th} spell, i.e. (T_{1k}, T_{2k}), is generated by a unique copula $C_D(\cdot, \boldsymbol{\theta}_D)$, while the joint distribution between all latent durations for all risks and spells is generated by the unique copula $C_M(\cdot, \boldsymbol{\theta}_M)$ which links the joint survival function of two competing latent durations for the K spells. The nested copula structure for $K = 2$ is therefore

$$\begin{aligned}&\Pr(T_{11} > t_{11}, T_{21} > t_{21}, T_{12} > t_{12}, T_{22} > t_{22}|\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\theta}_D, \boldsymbol{\theta}_M)\\ &= C_M\{C_D[S_{11}(t_{11}; \boldsymbol{x}_1, \boldsymbol{\beta}_1), S_{21}(t_{21}; \boldsymbol{x}_1, \boldsymbol{\beta}_1); \boldsymbol{\theta}_D], C_D[S_{12}(t_{12}; \boldsymbol{x}_2, \boldsymbol{\beta}_2), S_{22}(t_{22}; \boldsymbol{x}_2, \boldsymbol{\beta}_2); \boldsymbol{\theta}_D]; \boldsymbol{\theta}_M\}\end{aligned} \quad (3)$$

The dimension of the copula parameters $\boldsymbol{\theta}_D$ and $\boldsymbol{\theta}_M$ is copula specific and typically 1 or 2. Without further restrictions on C_D, the dependence structure between competing risks, S_{jk} are not identified nonparametrically for $j = 1, 2$ (Lo et al., 2020). While C_M is identified, C_D cannot or can only be weakly identified (Emura et al., 2020). We will only use Archimedean copulas C_M and C_D in this paper for reasons of practicability, even though the model is compatible with more general copula structures to some extent. For example, simulating the model without nested Archimedean copula structure is a nearly impossible task. In this paper we specify Q_j and C_M as this does not require non-testable restrictions on C_D and S_{jk}. It is therefore the equivalent of only studying Q_j in the single spell framework as considered by Bakoyannis and Touloumi (2012). We can therefore leave C_D and $\boldsymbol{\theta}_D$ unspecified.

Proposition 1 *$\boldsymbol{\beta}$ and $\boldsymbol{\theta}_M$ in models (1) and (3) are identifiable given observable $(\boldsymbol{T}, \boldsymbol{\delta})$.*

This follows directly from the fact that Q_j for $j = 1, 2$ are identifiable from observed data. Identifiability of $\boldsymbol{\theta}_M$ follows from Proposition 1 in Lo et al. (2020) which can be carried over to the model in this paper. A proof of Proposition 1 is therefore omitted.

2.1 Estimation

Estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_M$ is by maximum likelihood and bases on the joint distribution of observed durations and observed failure types. For a random sample of units i of size n, each i has K_i spells, where K_i is a positive integer that is small in typical applications. Let $\{t_{ijk}, j = 1, 2; k = 1, \ldots, K_i\}$ be $i = 1, \ldots, n$ independent realisations from model (3). Define $t_{ik} = \min\{t_{i1k}, t_{i2k}\}$, and $\delta_{ik} = \text{argmin}\{j : t_{ijk}\}$. Due to independent right censoring, either (t_{ik}, δ_{ik}) or c_{ik} is observed for all i and k, where censoring time C_k is independent of (T_k, Δ_k). Let $y_{ik} = \min\{t_{ik}, c_{ik}\}$, and $\eta_{ik} = \delta_{ik} \times \mathbb{1}(t_{ik} < c_{ik})$. Hence, $\eta_{ik} = 0$ if the k-th spell of unit i is censored, and $\eta_{ik} = \delta_{ik}$ otherwise. Let $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{iK_i})$, $\boldsymbol{y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n)$, $\boldsymbol{\eta}_i = (\eta_{i1}, \cdots, \eta_{iK_i})$, and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_n)$. The observed data are therefore $(\boldsymbol{y}, \boldsymbol{\eta}, \boldsymbol{x})$. The unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_M$ in models (1) and (3) can be estimated by maximum likelihood. The sample likelihood is

$$
\begin{aligned}
& L(\boldsymbol{\beta}, \boldsymbol{\theta}_M; \boldsymbol{y}, \boldsymbol{\eta}, \boldsymbol{x}) \\
= & \prod_{i=1}^{n} L_i(\boldsymbol{\beta}, \boldsymbol{\theta}_M; \boldsymbol{y}_i, \boldsymbol{\eta}_i, \boldsymbol{x}_i) \\
= & \prod_{i=1}^{n} (-1)^{K_i} \frac{\partial^{K_i} C_M\{S(y_1; \boldsymbol{x}_1, \boldsymbol{\beta}_1), \cdots, S(y_{K_i}; \boldsymbol{x}_{K_i}, \boldsymbol{\beta}_{K_i}); \boldsymbol{\theta}_M\}}{\partial y_1 \cdots \partial y_{K_i}} \Bigg|_{(y_1, \cdots, y_{K_i}, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{K_i}) = (y_{i1}, \cdots, y_{iK_i}, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iK_i})} \\
= & \prod_{i=1}^{n} \Bigg[\frac{\partial^{K_i} C_M\{S(y_1; \boldsymbol{x}_1 \cdot, \boldsymbol{\beta}_1), \cdots, S(y_{K_i}; \boldsymbol{x}_{K_i}, \boldsymbol{\beta}_{K_i}); \boldsymbol{\theta}_M\}}{\partial S(y_1; \boldsymbol{x}_1, \boldsymbol{\beta}_1) \cdots \partial S(y_{K_i}; \boldsymbol{x}_{K_i}, \boldsymbol{\beta}_{K_i})} \Bigg|_{(y_1, \cdots, y_{K_i}, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{K_i}) = (y_{i1}, \cdots, y_{iK_i}, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iK_i})} \\
& \times \prod_{k=1}^{K_i} \left\{ q_1(y_{ik}; \boldsymbol{x}_{ik}, \boldsymbol{\beta}_{1k})^{(2-\eta_{ik})} q_2(y_{ik}; \boldsymbol{x}_{ik}, \boldsymbol{\beta}_{2k})^{(\eta_{ik}-1)} \right\}^{\mathbf{1}\{\eta_{ik}>0\}} S(y_{ik}; \boldsymbol{x}_{ik}, \boldsymbol{\beta}_k)^{\mathbf{1}\{\eta_{ik}=0\}} \Bigg]
\end{aligned} \tag{4}
$$

with $S(y_k; \boldsymbol{x}_k, \boldsymbol{\beta}_k) = C_D[S_{1k}(y_k; \boldsymbol{x}_k, \boldsymbol{\beta}_k), S_{2k}(y_k; \boldsymbol{x}_k, \boldsymbol{\beta}_k); \boldsymbol{\theta}_D]$ and $q_j(t_k; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) = \partial Q_j(t_k; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk})/\partial t_k$. Because C_D and θ_D are unknown, these components are partialed out of the likelihood by modelling the distribution of observable y_k. After taking the logarithm, the log likelihood $\ln L_i$ for unit i consists of two additive parts

$$
\ln L_i = \ln L_{1i}(\boldsymbol{\beta}, \boldsymbol{\theta}_M; \boldsymbol{y}_i, \boldsymbol{x}_i) + \ln L_{2i}(\boldsymbol{\beta}; \boldsymbol{y}_i, \boldsymbol{\eta}_i, \boldsymbol{x}_i)
$$

with

$$
\begin{aligned}
& \ln L_{1i}(\boldsymbol{\beta}, \boldsymbol{\theta}_M; \boldsymbol{y}_i, \boldsymbol{x}_i) \\
= & \ln \left\{ \frac{\partial^{K_i} C_M\{S(y_1; \boldsymbol{x}_1, \boldsymbol{\beta}_1), \cdots, S(y_{K_i}; \boldsymbol{x}_{K_i}, \boldsymbol{\beta}_{K_i}); \boldsymbol{\theta}_M\}}{\partial S(y_1; \boldsymbol{x}_1, \boldsymbol{\beta}_1) \cdots \partial S(y_{K_i}; \boldsymbol{x}_{K_i}, \boldsymbol{\beta}_{K_i})} \Bigg| \right\}_{(y_1, \cdots, y_{K_i}, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{K_i}) = (y_{i1}, \cdots, y_{iK_i}, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iK_i})}
\end{aligned} \tag{5}
$$

and

$$
\begin{aligned}
& \ln L_{2i}(\boldsymbol{\beta}; \boldsymbol{y}_i, \boldsymbol{\eta}_i, \boldsymbol{x}_i) \\
= & \sum_{k=1}^{K_i} \Big\{ (2-\eta_{ik}) \mathbb{1}\{\eta_{ik} > 0\} \ln q_1(y_{ik}; \boldsymbol{x}_{ik}, \boldsymbol{\beta}_{1k}) + (\eta_{ik} - 1) \mathbb{1}\{\eta_{ik} > 0\} \ln q_2(y_{ik}; \boldsymbol{x}_{ik}, \boldsymbol{\beta}_{2k}) \\
& + \mathbb{1}\{\eta_{ik} = 0\} \ln S(y_{ik}; \boldsymbol{x}_{ik}, \boldsymbol{\beta}_k) \Big\}. \qquad (6)
\end{aligned}
$$

The model can be estimated by means of different approaches. L_{2i} is a function of $\boldsymbol{\beta}$ only. Therefore, $\boldsymbol{\beta}$ can be consistently estimated by maximising $\ln L_{2i}$ using partial likelihood. By only maximising $\ln L_{2i}$, the dependence structure between spells is ignored. This comes down to partial maximum likelihood by pooling all spells from all units and applying a usual single spell model. But $\ln L_{2i}$ only corresponds to the full likelihood if the multiple spells are independent, as $\ln L_{1i}$ is zero for all i in the case of independence copula. In presence of spell dependencies, this approach is still consistent but inefficient and leads the information matrix equality to fail. In consequence, the usual MLE inference is invalid. Maximising $\ln L_{1i}$ and $\ln L_{2i}$ jointly is efficient and the usual MLE inference is valid. In the case the full likelihood is too complicated, it is also possible to pre-estimate $\boldsymbol{\beta}$ by maximising $\ln L_{2i}$ in a first step and then, in a second step, maximise $\ln L_{1i}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}_M; \boldsymbol{y}_i, \boldsymbol{x}_i)$ in $\boldsymbol{\theta}_M$. This is also less efficient than one-step estimation but gives an estimate for $\boldsymbol{\theta}_M$. Joe (2005) shows that there is some loss in efficiency in two-step estimation of copula models but the loss is not large. In our simulations we compare the performance of pooled and one-step estimation to check whether ignoring spell dependencies is expected to lead to considerable loss in efficiency.

2.2 Choices for Q_j and C_M

There is a wide range of choices for Q_j and C_M that are compatible with models (1) and (3). In the following, we list some popular examples. Q_j can be directly modelled or is implied by models for the cause-specific or subdistribution hazards. In addition to parametric models with a small number of shape parameters, there are more flexible choices such as piecewise constant and semiparametric models. The cause-specific hazard is

$$
\lambda_j(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) = \lim_{\epsilon \to 0} \Pr(t \le T_k < t + \epsilon, \Delta_k = j | T_k \ge t, \boldsymbol{x}_k, \boldsymbol{\beta}_{jk})/\epsilon. \qquad (7)
$$

The cumulative incidence curves and overall survival function can be computed from the cause-specific hazards as follows:

$$
\begin{aligned}
Q_j(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) &= \int_0^t \lambda_j(s; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) \exp\left[-\int_0^s \lambda_1(u; \boldsymbol{x}_k, \boldsymbol{\beta}_{1k}) + \lambda_2(u; \boldsymbol{x}_k, \boldsymbol{\beta}_{2k}) du \right] ds, \qquad (8) \\
S(t; \boldsymbol{x}_k, \boldsymbol{\beta}_k) &= \exp\left[-\int_0^t \lambda_1(s; \boldsymbol{x}_k, \boldsymbol{\beta}_{1k}) + \lambda_2(s; \boldsymbol{x}_k, \boldsymbol{\beta}_{2k}) ds \right]. \qquad (9)
\end{aligned}
$$

By specifying a functional form for λ_j, (8) and (9) can be substituted into the likelihood function (4) to estimate parameters $\boldsymbol{\beta}$ without directly specifying or modelling Q_j. We suggest the following generalisation of popular single spell parametric models for λ_j that additionally allow parameters to be spell specific:

(i) Log-normal accelerated failure time model (LNAFT) with parameters $\boldsymbol{\beta}_{jk} = (\nu_{jk}, \rho_{jk}, \boldsymbol{\gamma}_{jk})$, where $\nu_{jk} \in \mathbb{R}_+$ and $\rho_{jk} \in \mathbb{R}$ are the risk and spell specific shape or nuisance parameters and $\boldsymbol{\gamma}_{jk}$ are risk and spell specific parameters on the covariates:

$$\lambda_j(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) = \frac{\frac{1}{\nu_{jk} t} \phi\left(\frac{\ln t - \rho_{jk} - \boldsymbol{x}_k' \gamma_{jk}}{\nu_{jk}}\right)}{1 - \Phi\left(\frac{\ln t - \rho_{jk} - \boldsymbol{x}_k' \gamma_{jk}}{\nu_{jk}}\right)}; \tag{10}$$

(ii) Log-logistic proportional odds model (LLPOM) with parameters $\boldsymbol{\beta}_{jk} = (\nu_{jk}, \rho_{jk}, \boldsymbol{\gamma}_{jk})$, where $\nu_{jk} \in \mathbb{R}$ and $\rho_{jk} \in \mathbb{R}_+$ are again the risk and spell specific shape parameters and $\boldsymbol{\gamma}_{jk}$ are risk and spell specific parameters on the covariates:

$$\lambda_j(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) = \frac{\nu_{jk} \rho_{jk} (\nu_{jk} t)^{(\rho_{jk} - 1)} \exp(\boldsymbol{x}_k' \boldsymbol{\gamma}_{jk})}{1 + (\nu_{jk} t)^{\rho_{jk}} \exp(\boldsymbol{x}_k' \boldsymbol{\gamma}_{jk})}; \tag{11}$$

(iii) Odd-rate transformation model with Gompertz cause-specific baseline hazard (ORGOM), with parameters $\boldsymbol{\beta}_{jk} = (\nu_{jk}, \rho_{jk}, \zeta_{jk}, \boldsymbol{\gamma}_{jk})$, where $\nu_{jk} \in \mathbb{R}_+$, $\rho_{jk} \in \mathbb{R}_+$, and $\zeta_{jk} \in \mathbb{R}$ are the risk and spell specific nuisance parameters and and $\boldsymbol{\gamma}_{jk}$ are risk and spell specific parameters on the covariates:

$$\lambda_j(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) = \frac{\nu_{jk} \exp(\boldsymbol{x}_k' \boldsymbol{\gamma}_{jk}) \exp(\rho_{jk} t)}{1 + \zeta_{jk} \nu_{jk} \exp(\boldsymbol{x}_k' \boldsymbol{\gamma}_{jk})(\exp(\rho_{jk}) - 1)/\rho_{jk}} \tag{12}$$

Another alternative is a proportional piecewise constant (PWCON) model for the cause-specific hazard functions (7). The idea is here to approximate an arbitrary unknown baseline hazard $\lambda_{j0}(t; ...)$ as a sequence of constants. The smaller the distance between the grid points, the more accurate the approximation. Even though the model is parametric, it has a semiparametric Cox model nature if a fine enough approximation of the unknown baseline is chosen. The model in the single spell context has been considered by Lawless (2003), Craiu and Duchesne (2004), Craiu and Lee (2005) and Kyyrä (2009) but has not been equivalently popular as the piecewise constant model for the marginal hazard. We suggest a generalised version that allows for risk and spell specific components:

$$\lambda_j(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) = \lambda_{j0}(t; \boldsymbol{\alpha}_{jk}) \phi_j(\boldsymbol{x}_k, \boldsymbol{\gamma}_{jk}), \tag{13}$$

where $\phi_j(\boldsymbol{x}_k, \boldsymbol{\gamma}_{jk}) = \exp(\boldsymbol{x}_k' \boldsymbol{\gamma}_{jk})$ is typically assumed. The baseline cause-specific hazard function $\lambda_{j0}(t; \boldsymbol{\alpha}_{jk})$ is modelled as a series of P_j discontinuous horizontal lines:

$$\lambda_{j0}(t; \boldsymbol{\alpha}_{jk}) = \sum_{p=1}^{P_j} \alpha_{jk,p} \mathbb{I}\{a_{jk,p} \leq t < a_{jk,p+1}\}, \text{where } a_{jk,1} = 0, a_{jk,P_j+1} = \infty \tag{14}$$

with unknown parameters $\boldsymbol{\alpha}_{jk} = (P_j, \alpha_{jk,1}, \cdots, \alpha_{jk,P_j}, a_{jk,2}, \cdots, a_{jk,P_j})'$ and thus $\boldsymbol{\beta}_{jk} = (\boldsymbol{\alpha}_{jk}, \boldsymbol{\gamma}_{jk})$.

Due to the absence of a semiparametric Cox-type proportional hazard model for the cause-specific hazard, a semiparametric model for the subdistribution hazard is widely used in applied research. The model by Fine and Gray (1999) is among the most popular analysis models for competing risks models. The subdistribution hazard is

$$\lambda_j^s(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) = \lim_{\epsilon \to 0} \Pr(s \leq T_k < s + \epsilon, \Delta_k = j | T_k \geq s \vee (T_k < s \wedge \Delta_k \neq j)/\epsilon.$$

It has a different conditioning set than the cause-specific hazard and its interpretation is difficult (Bakoyannis and Touloumi, 2012). However, it is possible to obtain Q_j from λ_j^s through

$$Q_j(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) = 1 - \exp\{-\Lambda_{0j}^s(t) \exp(\boldsymbol{x}_k' \boldsymbol{\gamma}_{jk})\}, \tag{15}$$

with $\Lambda_{0j}^s(t)$ is the baseline cumulative subdistribution hazard

$$\Lambda_{0j}^s(t) = \int_0^t \lambda_{0j}^s(s) \, ds. \tag{16}$$

$\lambda_{0j}^s(t)$ is a nonparametric subdistribution baseline hazard and $\exp(\boldsymbol{x}_k' \boldsymbol{\gamma}_{jk})$ is the proportional shifter of the baseline hazard that depends on covariate values. The single spell version of this model is implemented in main statistical packages such as R and STATA. One can use this implementation to obtain estimates for $\Lambda_{0j}^s(t)$ and $\boldsymbol{\beta}$ and therefore Q_j. $\boldsymbol{\theta}_M$ can then be estimated in a second step by maximising likelihood (5) for given $\hat{\boldsymbol{\gamma}}_{jk}$ and $\hat{Q}_j$. For comparison with the direct models of the cause-specific hazard, we consider this model in our application. In particular, we consider whether it matters for empirical results if one specifies the cause-specific or the subdistribution hazard.

In what follows, we discuss the choice of copula. One popular choice for $C_M(\ldots; \boldsymbol{\theta}_M)$ in model (3) is the family of Archimedean copulas with symmetry property in the sense that all arguments in the copula function can be interchanged. As a result, the inter-spell dependencies are identical for any pair of spells. More details about copulas can be found in Nelsen (2006) and Trivedi and Zimmer (2005). In our numerical analysis we only consider one parameter copulas and we write $\boldsymbol{\theta}_M = \theta$ for simplicity. In this case, there is a 1-1 link between the parameter and Kendall-$\tau \in [-1, 1]$, the measure of rank correlation. Given that Kendall-τ has a clearer interpretation, we consider it in our numerical analysis, where we work with the following popular Archimedean copulas:

(i) Clayton copula, with $\theta > 0$:

$$C(s_1, \cdots, s_k; \theta) = (s_1^{-\theta} + \cdots + s_k^{-\theta} - (k-1))^{-1/\theta}. \tag{17}$$

Kendall-$\tau = \theta/(\theta+2)$. Thus, $\tau > 0$ and in a limiting case where $\theta \to 0$, the Clayton copula becomes an independence copula with $\tau \to 0$. The Clayton copula has asymmetric tail dependence with strong dependence in the left tail, while weak dependence in the right tail. The Clayton copula is therefore best suited for applications in which repeated durations are likely to experience low values together, while there is no relationship for long durations.

(ii) Frank copula, with $\theta \in \mathbb{R}\backslash\{0\}$:

$$C(s_1, \cdots, s_k; \theta) = -\frac{1}{\theta} \ln \left[\frac{(\exp(-\theta s_1) - 1) \cdots (\exp(-\theta s_k) - 1)}{(\exp(-\theta) - 1)^{k-1}} + 1 \right]. \tag{18}$$

For $\theta = 0$, it reduces to the independence copula with Kendall-$\tau = 0$. The tail dependence of the Frank copula is symmetric with stronger dependence in the centre and weaker dependence at the tails. The Frank copula is therefore suitable in applications in which tail dependence is weaker.

(iii) Gumbel copula with $\theta \geq 1$:

$$C(s_1, \cdots, s_k; \theta) = \exp \left[- \left\{ (-\ln s_1)^{\theta} + \cdots + (-\ln s_k)^{\theta} \right\}^{1/\theta} \right]. \tag{19}$$

Kendall-$\tau = 1 - 1/\theta$. Thus, $\tau > 0$. The Gumbel copula has asymmetric tail dependence, with stronger right tail dependence and weaker left tail dependence.

The choice of copula is ideally motivated by some knowledge or hypothesis about the dependence structure. The evidence based on simulations in Lo et al. (2020) suggests that the choice of Archimedean copula plays only a limited role for the estimated marginal survival curves. A generalisation of the dependence structure to asymmetric dependencies that allow dependencies to be pairwise different between multiple spells has been suggested in Lo et al. (2020) but more general identification results are still to be developed.

2.3 Model selection test

In the case Q_j or C_M are unknown in an application, one can use the following model selection test to reject models. The classical likelihood ratio tests cannot be conducted because the different specifications of Q_j or C_M lead to non-nested models. An inference approach for non-nested model selection tests has been introduced by Vuong (1989). The test statistic is

$$V = \frac{n^{-1} \sum_{i=1}^{n} [\ln L_i^{**}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}_i, \boldsymbol{\eta}_i, \boldsymbol{x}_i) - \ln L_i^{*}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}_i, \boldsymbol{\eta}_i, \boldsymbol{x}_i)]}{\{n^{-1} \sum_{i=1}^{n} [\ln L_i^{**}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}_i, \boldsymbol{\eta}_i, \boldsymbol{x}_i) - \ln L_i^{*}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}_i, \boldsymbol{\eta}_i, \boldsymbol{x}_i)]^2\}^{1/2} / \sqrt{n}}, \tag{20}$$

with L^{**} is the likelihood function for the tested model while L^* is that for the alternative model. V is asymptotically standard normal distributed under the null hypothesis that $E(\ln L^*) = E(\ln L^{**})$. It is important to mention that the test bases on a quasi-maximum likelihood approach and for the statistic to be normal it is required that both models compared are incorrectly specified. It is therefore not possible to find the true model with this test but only to reject models with a greater degree of misspecification. In this paper we apply this test for the first time in the context of copula duration models as a model selection tool.

2.4 Interpretability and Partial Effects

Depending on the chosen model for λ_j, λ_j^s and C_M, the parameters obtained by maximising the likelihood (4) have different interpretations. It is therefore indispensable to consider unified measures to compare results across models. Besides from comparing the estimated Q_j, we also consider by how much the estimated partial covariate effects differ. This is by how much the cumulative incidence increases or decreases in response to a change in one variate x_l in spell k holding all other variables constant, i.e. $Q'_{jkl}(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) = \partial Q_j(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk})/\partial x_{kl}$ for a continuous x_l. These partial effects are of key interest in applications but unfortunately, they are complicated and not just the relevant parameter β_{jkl} (Kyyrä, 2009). In fact, not even the sign of β_{jkl} determines the direction of the partial effect and the sign of the partial effect can change with t.

Proposition 2 *It is possible that there exist* $t_1 \neq t_2$ *such that* $signQ'_{jkl}(t_1; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) \neq signQ'_{jkl}(t_2; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk})$ *for some* $j = 1, 2,\ k = 1, \dots, K$ *and* $l = 1, \dots, L_k$.

For a proof of the proposition see the Appendix. Whether there is a change in the sign of the covariate effect as duration increases depends on the chosen model for Q_j, its parameters and the covariate. The result complements the observation by Kyrrä (2009) that the direction of the partial effect is not unique for all t when a PWCON model for the cause-specific hazard functions is used. There are, however, models for which the direction of the partial effect is the same for all t. A prominent example is the semiparametric model for the subdistribution hazard by Fine and Gray (1999) for which

$$Q'_{jkl}(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) = (Q_j(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) - 1) \ln(1 - Q_j(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}))\beta_{jkl}.$$

Thus, in this model the magnitude of the partial effect varies with t but its direction is determined by the sign of β_{jkl} for all t. This model therefore restricts more strongly the partial effects than flexible proportional hazard models for the cause-specific hazard, including the PWCON model. Our data analysis in Section 4 presents examples for this.

3 Simulations

We conduct a simulation study to investigate the finite sample performance of the models described in Section 2. Previous simulations in Lo et al. (2020) are for a simplified model without regressors and spell-varying model components, and are limited to parametric models. Monte Carlo evidences for parametric and semiparametric models with regressors can be found in Lo and Wilke (2014) but are restricted to the single spell model and therefore non-nested copula structure. In our simulations we consider a generalised simulation design with nested copula structure and spell-varying covariates $\boldsymbol{x}_k$ and their parameters. The simulation procedure is described in the Supplementary Material S.I. In addition to parametric models for the cause-specific hazard, we study the behaviour of a flexible piecewise constant cause-specific hazard model and compare it to the well-known semiparametric proportional hazard model for the subdistribution (Fine and Gray, 1999). The latter allows us to investigate whether it matters in practice which type of hazard function is being modelled and estimated. For the cumulative incidences, we estimate Q_j and partial covariate effects on Q_j. For the copula parameters, we estimate the Kendall-τ for the copula C_M, which is written as τ_M for simplicity. We do not estimate the Kendall-τ for the copula C_D, but this parameter is involved in the data generating process (DGP) and thus is used to simulate the model, so we write it as τ_D to distinguish it from τ_M.

We draw a random sample with $n = 2,000$ units, where 50% have $K_i = 1$, 25% have $K_i = 2$ and the rest have $K_i = 3$. K_i is random for all i. The pooled sample therefore consists of $m = 3,500$ spells. In all simulation designs, we draw data from a known two-risks model with log-normal accelerated failure time cause-specific hazards and the Frank copula for both C_D and C_M. We simulate two binary $\boldsymbol{x}_k = (x_{k1}, x_{k2})$, where $x_{kl} = \{0, 1\}$ with $Pr(x_{kl} = 1) = 0.5$ for $k = 1, 2, 3$ and $l = 1, 2$. Among them, x_{k1} does not change across spells, while x_{k2} can be different for each spell, i.e. $x_{11} = x_{21} = x_{31}$ and this restriction does not hold for x_{k2}. We sample 500 times. The parameters of the data generating process are given in Table 1.

We assess the performance of the estimates for τ_M, Q_j and partial covariate effects on Q_j by computing their (average) squared bias and (average) mean squared error. For the partial covariate effects, we consider the partial effect of a discrete change in x_{k1} and x_{k2} on Q_j, where

$$\begin{aligned} dQ_jd_1(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) &= Q_j(t; x_{k1} = 1, x_{k2} = 0, \boldsymbol{\beta}_{jk}) - Q_j(t; x_{k1} = 0, x_{k2} = 0, \boldsymbol{\beta}_{jk}), \\ dQ_jd_2(t; \boldsymbol{x}_k, \boldsymbol{\beta}_{jk}) &= Q_j(t; x_{k1} = 0, x_{k2} = 1, \boldsymbol{\beta}_{jk}) - Q_j(t; x_{k1} = 0, x_{k2} = 0, \boldsymbol{\beta}_{jk}). \end{aligned}$$

While the squared bias and the mean squared error are obvious to compute for $\hat{\tau}_M$, the measures for $\hat{Q}_j$ and partial effects vary in t. For the latter we therefore consider the average squared bias

(ASB) and average mean squared error (AMSE) over t for the reference unit $\boldsymbol{x}_k = (0,0)$ which are

$$
\begin{aligned}
ASB(\hat{Q}_j) &= \frac{1}{G}\sum_{g=1}^{G}\Big(E[\hat{Q}_j(t_g;\boldsymbol{x}_k,\hat{\boldsymbol{\beta}}_{jk}) - Q_j(t_g;\boldsymbol{x}_k,\boldsymbol{\beta}_{jk})]\Big)^2, \\
AMSE(\hat{Q}_j) &= \frac{1}{G}\sum_{g=1}^{G} E[(\hat{Q}_j(t_g;\boldsymbol{x}_k,\hat{\boldsymbol{\beta}}_{jk}) - Q_j(t_g;\boldsymbol{x}_k,\boldsymbol{\beta}_{jk}))^2]
\end{aligned}
$$

for risks $j = 1,2$ and spells $k = 1,2,3$, where $t_g = t_1,\ldots,t_G$ are grid points on the support of T with $G = 3,500$. We estimate the expected values by taking the average over the resulting 500 estimates. As Q_j partly varies across models due to different data generating processes, the ASB and AMSE should be related to their magnitude for comparability. We therefore also consider a relative AMSE (RAMSE) which has a percentage interpretation. It is obtained by dividing AMSE($\hat{Q}_j$) by $\bar{Q}_j = \frac{1}{G}\sum_{g=1}^{G} E[Q_j(t_g;\boldsymbol{x}_k,\boldsymbol{\beta}_{jk})]$. It is zero if the AMSE was zero and it is one if the AMSE had the same value as $\bar{Q}_j$. For the partial covariate effects, we replace $Q_j(t_g;\boldsymbol{x}_k,\boldsymbol{\beta}_{jk})$ with $dQ_j d_l(t_g;\boldsymbol{x}_k,\boldsymbol{\beta}_{jk})$.

In what follows, we examine model performance when, first, the cumulative incidence is misspecified, second, C_M is misspecified, and third, coefficients of covariates vary across spells. For now, we study the performance of the model when the cumulative incidence is misspecified when assuming the correct C_M. The models for the cause-specific hazard λ_j comprise the log-normal accelerated failure time model (LNAFT), the log-logistic proportional odds model (LLPOM), the odd-rate transformation model with Gompertz cause-specific baseline hazard (ORGOM), and the model with piecewise constant cause-specific hazard (PWCON). This is complemented by the semiparametric proportional hazard model for the subdistribution hazard (semiparametric PH). We present the results for 6 different estimated models. The estimated models are as follows: Model 1 is correctly specified. Models 2-5 have misspecified Q_j. Model 4 uses the flexible piecewise constant cause-specific hazard model with 15 intervals for each risk. Model 5 is a two-step approach as described in Section 2, where the first step is a partial MLE (PMLE) with Q_j implied by the semiparametric proportional hazard model (Fine and Gray, 1999), and the second step is estimating τ_M by maximising likelihood (5) given $\hat{Q}_j$. Model 6 is a partial MLE with the correct model for Q_j that only maximises (6) and therefore does not estimate τ_M. It is to analyse by how much efficiency is affected when a part of the likelihood is ignored in the estimation. In this simulation design the parameters do not change across spells, i.e. $\boldsymbol{\beta}_{j1} = \boldsymbol{\beta}_{j2} = \boldsymbol{\beta}_{j3}$ and for this reason we remove k in this setup.

Table 1 shows the results for all models. Models 1-3, the first three parametric models, including the correctly specified model, do a good job in estimating cumulative incidences, partial effects, and the dependence structure. While all measures are low, the correctly specified model performs best as expected. The similar performance of the three models can be explained by the fact that the underlying true parametric functions are smooth and regular and can be well approximated by the misspecified models. The performance of Model 4 is worse than Models 1-3 but not far off. Given that Model 4 works under very mild parametric restrictions, it is able to approximate hazards without knowing their exact functional forms, and provide an accurate estimate for the degree of spell dependence. Model 5 also provides an accurate estimate for the degree of spell dependence. However, despite that it is even more flexible than Model 4, it results in larger ASB and AMSE for $\hat{Q}_j$ and $\widehat{dQ_j d_l}$. This can be explained by the fact that the assumed restrictions on the role of covariates, i.e. proportional hazards, are for the subdistribution hazards and not cause-specific hazards. Making the same restriction on the two different hazards is not compatible with the same Q_j and in our simulation design this makes the results for Model 5 worse. Figure 1 presents the mean of the estimated functionals for Models 1-5 and compares them to the data generating process. For models 1-5 we plot the average of estimates over the 500 simulations against time. We find that Model 1 with the correct specification overlaps with the true model for all cumulative incidences and partial effects. Models 2-3 follow the true model closely, but not as good as Model 1. Models 4-5 follow the shape of the true model to some extent, but their estimates are further away from the true values compared to Models 1-3. Inspecting in more detail, Model 4 and Model 5 are close to each other for shorter durations, and for longer durations Model 4 is closer to the true model than Model 5, which makes Model 4 usually outperform Model 5. In fact, we can largely control the performance of the piecewise constant model by adjusting the intervals. To show this more clearly, we estimate three piecewise constant models using different number of intervals, which is 7, 12, 15 for M1, M2 and M3, respectively. We show the bias and mean squared error related measures for the three models in Table S1, and the comparison of the three models with the true data generating process and semiparametric PH model in Figure S1 in the Supplementary Material S.II. The performance increases a lot from PWCON M1 to PWCON M2, yet not much from PWCON M2 to PWCON M3. In general, we find that the finer the intervals, the more closely it follows the true model. In the end, we choose to report PWCON M3 as Model 4 in Table 1.

Model 6 and the first step of Model 5 adopt a partial MLE approach by maximising likelihood (6) and ignoring the dependence structure. With correctly specified cumulative incidence, Model 6 shows a negligible decrease in efficiency compared with Model 1 that uses the full MLE. We further add Model 7 in Table S2 in the Supplementary Material S.II, which is a partial MLE for the PWCON model. It also shows only a small decrease in efficiency compared with Model 4. These results suggest that the MSE disadvantage of Model 5 is not due to its partial nature. We investigate the reason for the MSE disadvantage of the semiparametric PH model by simulating the model without regressors (i.e. all $\boldsymbol{\gamma}$s are zero). Results are presented in Table S3 in the

Table 1: Bias and Mean Squared Error related measures for various models with $n = 2000$ and $m = 3500$ for 500 simulated samples. Assumed model features correct unless otherwise stated.

DGP: Frank, LNAFT, $\boldsymbol{\nu} = (2, 1.5)$, $\boldsymbol{\rho} = (0.5, 2)$, $\boldsymbol{\gamma}_1 = (-2, 2)$, $\boldsymbol{\gamma}_2 = (4, -4)$, $\tau_M = \tau_D = 0.3$									
Model		1			2			3	
Assumed		correct			LLPOM			ORGOM	
	(A)SB	(A)MSE	RAMSE	(A)SB	(A)MSE	RAMSE	(A)SB	(A)MSE	RAMSE
$\hat{\tau}_M$	0.0000	0.0002		0.0000	0.0003		0.0000	0.0002	
$\hat{Q}_1$	0.0000	0.0001	0.04%	0.0001	0.0002	0.07%	0.0001	0.0002	0.07%
$\widehat{dQ_1 d_1}$	0.0000	0.0001	0.04%	0.0002	0.0003	0.10%	0.0001	0.0002	0.08%
$\widehat{dQ_1 d_2}$	0.0000	0.0002	0.06%	0.0003	0.0005	0.16%	0.0004	0.0006	0.21%
$\hat{Q}_2$	0.0000	0.0000	0.07%	0.0000	0.0001	0.09%	0.0001	0.0001	0.22%
$\widehat{dQ_2 d_1}$	0.0000	0.0001	0.02%	0.0001	0.0002	0.04%	0.0001	0.0003	0.04%
$\widehat{dQ_2 d_2}$	0.0000	0.0000	0.07%	0.0000	0.0001	0.09%	0.0001	0.0001	0.21%
Model		4			5			6	
Assumed		PWCON			Semiparametric PH			PMLE	
	(A)SB	(A)MSE	RAMSE	(A)SB	(A)MSE	RAMSE	(A)SB	(A)MSE	RAMSE
$\hat{\tau}_M$	0.0001	0.0004		0.0001	0.0005				
$\hat{Q}_1$	0.0005	0.0006	0.21%	0.0042	0.0043	1.45%	0.0000	0.0001	0.04%
$\widehat{dQ_1 d_1}$	0.0019	0.0020	0.72%	0.0150	0.0151	5.43%	0.0000	0.0001	0.04%
$\widehat{dQ_1 d_2}$	0.0036	0.0038	1.29%	0.0101	0.0102	3.42%	0.0000	0.0002	0.07%
$\hat{Q}_2$	0.0006	0.0006	0.98%	0.0009	0.0009	1.41%	0.0000	0.0000	0.07%
$\widehat{dQ_2 d_1}$	0.0043	0.0045	0.74%	0.0033	0.0036	0.59%	0.0000	0.0001	0.02%
$\widehat{dQ_2 d_2}$	0.0005	0.0005	0.83%	0.0007	0.0007	1.16%	0.0000	0.0000	0.07%

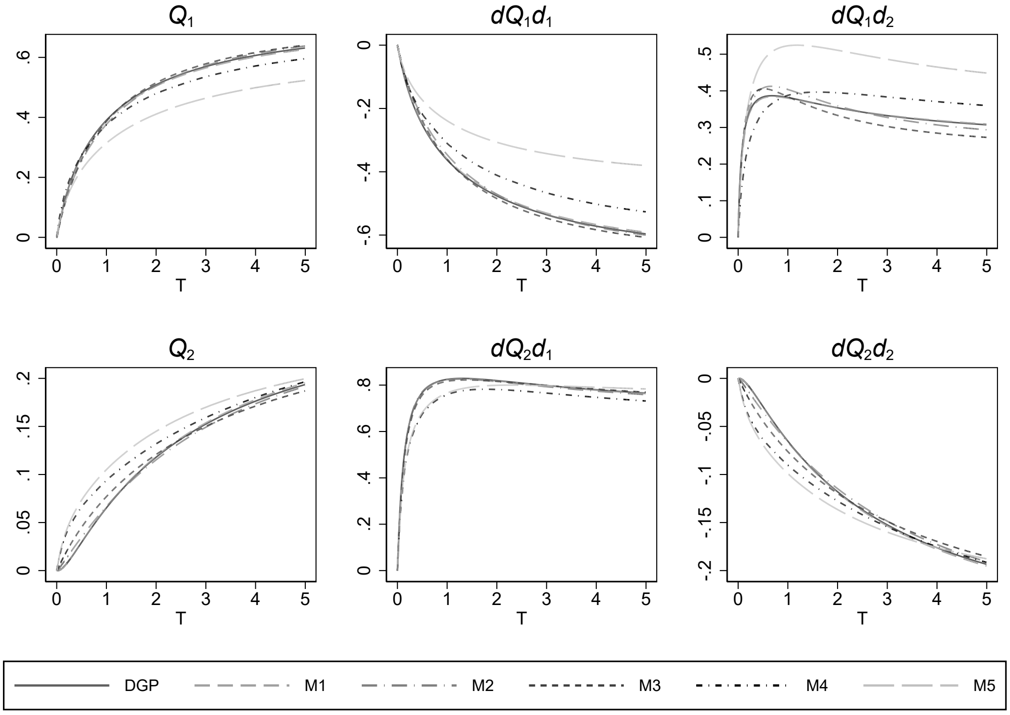

Figure 1: Comparison of Models 1–5 with the true DGP.

Supplementary Material S.II and show that the semiparametric PH model performs as good as the other parametric models. This provides evidence that the disadvantage of Model 5 in Table 1 comes from misspecification of the functional form for the role of regressors, leading to systematic bias of estimated Q_j and partial covariate effects. In addition, we find that the performance of PWCON is also better in Table S3 than in Table 1, suggesting that PWCON also suffers from misspecification of the functional form of the role of regressors, but the bias is smaller. The similar performance of Model 1 and Model 6 suggests that ignoring the dependence structure does not affect the estimation of cumulative incidences to a large extent. To see how robust this finding is, we simulate another model with a stronger dependence using a larger τ_M and τ_D for comparison. The estimation results in Table S4 in the Supplementary Material S.II show that differences in the performance of the two approaches are larger and the performance of the two approaches is worse when there is a strong dependence.

In the second step, we study the role of misspecifying the dependence structure C_M. In the context of our analysis this is assuming the Clayton or the Gumbel copula when the Frank copula is correct. The results are reported in Table S5 in the Supplementary Material S.II. Compared with Table 1, using the Gumbel copula leads to only small increases in ASB and AMSE for $\hat{\tau}_M$ and almost no change in $\hat{Q}_j$ and $\widehat{dQ_j d_l}$ for the first four parametric models, but for the semiparametric

PH model, the ASB and AMSE for $\hat{\tau}_M$ are larger. For the Clayton copula, the ASB and AMSE for $\hat{\tau}_M$ are larger than the Gumbel copula for all models, especially for the piecewise constant and the semiparametric PH model. However, the ASB and AMSE for $\hat{Q}_j$ and $\widehat{dQ_jd_l}$ are still almost unchanged. While misspecification of the cumulative incidence Q_j has less effects on $\hat{\tau}_M$ and more effects on $\hat{Q}_j$ and $\widehat{dQ_jd_l}$, misspecification of the Archimedean copula C_M does not affect the ASB and AMSE for $\hat{Q}_j$ and $\widehat{dQ_jd_l}$, but affects $\hat{\tau}_M$. In general, we find that compared with the Gumbel copula, the Clayton copula leads to larger ASB and AMSE for $\hat{\tau}_M$ for all models. In particular, when we use the semiparametric PH model and a misspecified copula - the Gumbel copula or the Clayton copula - the ASB and AMSE for $\hat{\tau}_M$ increase a lot. These results suggest that it is important not to choose the wrong copula in order to get a good estimate of τ_M.

We further apply the Vuong test for model selection, regarding the choice of copula and model for the cumulative incidence curve. Note that the Vuong test is for pairwise comparison of misspecified models, which means that we cannot include the correctly specified model in the test. However, by focusing on copula and cumulative incidence separately, we can still infer the best model out of the test results. That is, we can focus on the choice of copula when we use the same cumulative incidence, and focus on the choice of cumulative incidence when we use the same copula. We compute test statistics for pairwise comparisons between the misspecified parametric models, i.e. Models 2-4 in Table 1 and Models 1-4 and 6-9 in Table S5 in the Supplementary Material S.II. Each one of these 11 models is compared with the other 10, so in total we obtain 110 values of the rejection probability. Then, for each model, we calculate the average rejection probability across the 10 comparisons, so in total we obtain 11 values of average rejection probability. Figure 2 Panel A shows the average probability of rejecting each model. For the three copulas, the order of average rejection probability from smallest to largest is Frank, Gumbel, Clayton; and for the four cause-specific hazards, the order of average rejection probability from smallest to largest is LNAFT, LLPOM, ORGOM, PWCON. Given that the true model uses the Frank copula and LNAFT cause-specific hazard, the results suggest that the Vuong test does a good job in model selection. We further compute the ratio of sample log likelihood values for the two models in comparison. A ratio larger than 1 suggests that the model is a worse fit in the pairwise comparison. The larger the ratio, the worse the model performs relative to the other model. We plot the probability of rejecting a model against the sample log likelihood ratio in Figure 2 Panel B. It shows that the rejection probability increases with the sample log likelihood ratio, suggesting that a model with a worse fit is more likely to be rejected in the Vuong test. Overall, we find that the results in Table 1 and Table S5 correspond well to the results of the Vuong test. For example, the Clayton copula and piecewise constant model are shown to have larger ASB and AMSE among all copulas and cumulative incidence curves, and this is well captured by the average rejection probabilities of Clayton and PWCON in Figure 2 Panel A.

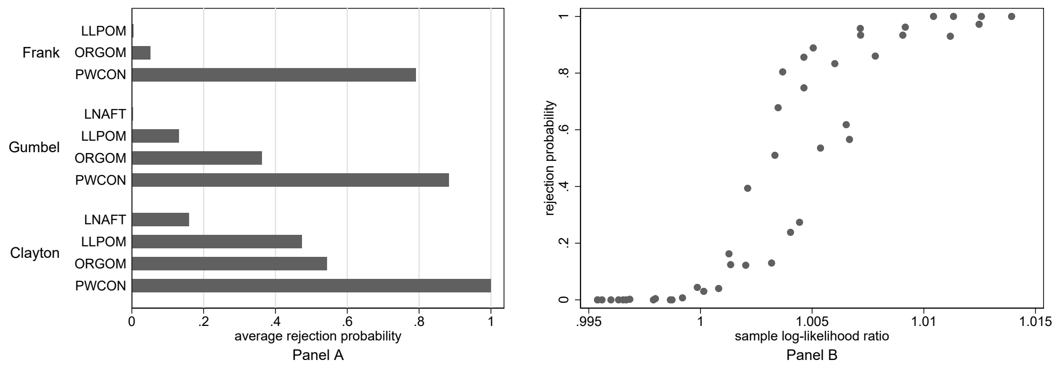

Figure 2: Rejection probability of the Vuong test.

In the last step, we study the performance of the models when coefficients of covariates change across spells. In particular, we set $\boldsymbol{\gamma}_{11} = (-1, 1)$, $\boldsymbol{\gamma}_{21} = (3, -3)$, $\boldsymbol{\gamma}_{12} = (-2, 2)$, $\boldsymbol{\gamma}_{22} = (4, -4)$, $\boldsymbol{\gamma}_{13} = (-3, 3)$, $\boldsymbol{\gamma}_{23} = (5, -5)$. In such a setting we allow partial covariate effects on cumulative incidence to differ across spells. We only study the role of misspecifying Q_j, as the role of misspecifying C_M will be similar to what we have shown before. We estimate Models 1, 2, 4, 5 of Table 1 and show the results in Table S6 in the Supplementary Material S.II. The estimated τ_M is similar for the four models, which replicates the findings in Table 1. Because we set the coefficients of the second spell to be the same as those in Table 1, the performance of $\hat{Q}_j$ and $\widehat{dQ_j d_l}$ for the four models is similar to Table 1 for the second spell. For the first and third spells, the first two parametric models still capture the spell-varying partial effects well. However, the ASB and AMSE of the partial effects for PWCON and the semiparmetric PH model vary across spells. In some cases, the ASB and AMSE are larger and in some cases are smaller. All in all, the comparison of the four models is very similar to Table 1, so the presence of spell-varying coefficients does not much affect previous findings.

4 Application

In this section we present a real world application with maternity leave periods to show the applicability of the model in Section 2. Given the ongoing demographic change with retrogressive birth rates in most industrialised countries, a better understanding of female behaviour after childbirth and the interplay with maternal labour supply is key to tackle this problem. Furthermore, empirically assessing maternity leave duration is of great importance since inactivity after childbirth has been found to be an important determinant of the gender wage gap (e.g. Waldfogel, 1997; Budig and England, 2001; Gangl and Ziefle, 2009; Schönberg and Ludsteck, 2007; Beblo et al., 2008).

It appears intuitive that post maternity leave labour market states and repeated maternity leave spells are not independent. It is also plausible that the covariate effects may vary across spells. In addition, biological constraints and incentives set by family policies (job protection periods, cash benefits during maternity leave etc.) make the distribution of maternity leave durations non smooth with mass points at certain times after childbirth. Therefore, allowing for nested dependencies between risks and spells and for a flexible shape of the hazard function is well justified for the analysis of maternity leave.

Duration models have been used for the analysis of maternity leave (Fitzenberger et al., 2016; Arntz et al., 2017; Rodrigues and Vergnat, 2019). However, these studies only use the first childbirth and abstract from maternity leave periods of any subsequent child. Our analysis design bases on Arntz et al. (2017), who study sorting into different exit states after the first maternity leave in Germany. In contrast, we adopt the nested copula model and consider maternity leaves after birth of the first three children. We use large scale biographical data from Germany (BASiD) which links administrative records from the German statutory pension insurance scheme (Rentenversicherung) and the Federal Employment Agency (Bundesagentur für Arbeit). In total, 579,000 individuals are included in the data which constitute a 1% random sample of around 96% of the German population. The dataset contains daily spell information about periods of employment, training, education and registered unemployment, demographic information (age, sex, region etc.) and information about birth dates of own children. For more details about the dataset see Hochfellner et al. (2012). We restrict our sample to females aged 18–45 who gave the first birth between 1985 and 2005 and who were dependently employed at the time of conception. This leaves us with 34,380 mothers. We construct maternity leave periods using the birth date of children and information about various other labour market states. Maternity leave is defined as any unobserved period after birth until exit into one of the four observable postmaternity states listed below. There is independent censoring at the end of the data in 2009. In addition, we censor durations after 50 months, since after this time only few transitions are observed.

For each female, we consider up to three maternity leave periods. Note that 97.3% of the mothers in our sample have at most three maternity leave periods and 44.4% have more than one. This means ignoring the 4^{th} or higher spells is only relevant for less than 3% of the mothers.

We use a model with 4 competing risks or post maternity leave states:

1. return to the same employer,

2. start a new job with a different employer,

3. have a next child (new maternity leave period),

4. enter a different state (education, training, self or minor employment, unemployment and other).

Table 2 shows exit frequencies by spell. It is apparent that the distribution of observed exit states varies strongly across spells. For example, while after the first spell, 19% of all mothers give birth to a next child and only 46% return to the same employer, after the second spell only 5% give birth and 50% return to the same employer.

Table 2: Number and share of spells by exist state.

	Spell 1		Spell 2		Spell 3	
	absolute	relative	absolute	relative	absolute	relative
Risk 1 - same employer	15,888	0.46	5,024	0.50	438	0.48
Risk 2 - new employer	3,814	0.11	1,240	0.12	123	0.14
Risk 3 - next child	6,569	0.19	544	0.05	32	0.04
Risk 4 - other	5,213	0.15	1,587	0.16	185	0.20
Censored	2,896	0.08	1,645	0.16	133	0.15
Sum	34,380	1.00	10,040	1.00	911	1.00

Notes: Based on German maternity leave data (BASiD).

We first estimate and compare the cumulative incidences for the log-normal accelerated failure time model (LNAFT), the log-logistic proportional odds model (LLPOM), the piecewise constant cause-specific hazard model (PWCON) and the semiparametric proportional hazard model (Fine and Gray, 1999). While the models with the first three cumulative incidences are estimated in one step by full ML, the model with semiparametric PH is estimated in two steps, with the first one being partial MLE as described in Section 3. The number and borders of the intervals are chosen to fit the patterns for observed durations as given in Figure 3, first row. For example, to well approximate the curve and mass points in panel (a), the interval borders for risk 1 are 0, 3, 5, 8, 11, 14, 20, 23, 26, 35, 38, 44, 50 which results in 12 intervals. Both, number and borders, can be different across risks. We use the Gumbel copula to model dependencies between spells because this copula is found to be most suitable as shown below. In the initial model we restrict parameters on regressors to not vary by spell, even though the regressors themselves are allowed to do so. Building on Arntz et al. (2017), the following categorical covariates are included: educational degree (no vocational degree, vocational degree, tertiary degree), living in former East Germany, age and employment status 10 months before birth (non-employed, part-time employed, full-time employed with earnings in the lower, middle or upper tercile). Applying the risk pooling approach (see Lo and Wilke, 2014), we estimate each model three times (for risks 1–3), each time for two risks where the first risk is the risk of interest and the second is a pool over all remaining risks. Since risk 4 consists of aggregated remaining states without clear interpretation, we do not consider explicit estimates for this risk.

Regression results for all models and risks are given in Table S7 in the Supplementary Material S.III. Broadly speaking, the estimated parameters indicate that being young, well-earning and/or from East Germany increases the probability of returning to the same employer, while being highly educated and/or not well-earning makes transitions into the second exit state, starting a job with a different employer, more likely. Young women from West Germany have a higher likelihood of having directly their next child, ceteris paribus. For all risks and models, the Kendall-τ varies between 0.15 and 0.27 with $\hat{\theta}$ always being highly significant. θ can be interpreted as a random frailty term due to the presence of omitted variables such as characteristics of the father, personal attitudes towards child-rearing and gender roles. For instance, being conservative is likely to prolong the duration before returning to work, and this for all spells. Similar to positive serial correlation in errors in unobserved effects panel models, the length of spells is expected to be strongly positively correlated across multiple occurrences. At the same time, being highly conservative might shorten the length of maternity leave before having a next child, again for all spells. It is therefore crucial that our model allows for different roles of the covariates across risks and for dependencies across spells.

Figure 3 visualises the share of females exiting into the respective state after a certain time after childbirth pooled over the first three spells (histograms, first row). The second row of the figure depicts the estimated cause-specific hazard functions for the various models and the third row the according cumulative incidence curves for the reference mother. For all characteristics, the reference is defined as the most frequently observed value at the first spell. This is West-German, aged between 24 and 28, has a vocational degree and works full-time with earnings in the second tercile of the income distribution. The histograms underline the existence and magnitude of mass points which correspond greatly to biological constraints and economic incentives due to the legal setup (see Arntz et al., 2017). For example, for risks one and two, a sizeable fraction of females enters the exit state after 36 months, which is exactly when the job protection period ends since 1992. The share of females having a next child drops sharply around the same time as many mothers attempt to have the next child within the job protection period of the previous child. Similar jumps can be found after 8 weeks, 6 months, 12 months and 24 months (when maternity leave and maternity benefits end). For the third risk, no transitions are observed before the biological minimum of ten months. The hazard curves portrayed in the second row illustrate the ability of the piecewise constant model to capture these mass points while the parametric models fail to do so. This is, though less visible, translated in a worse fit of the cumulative

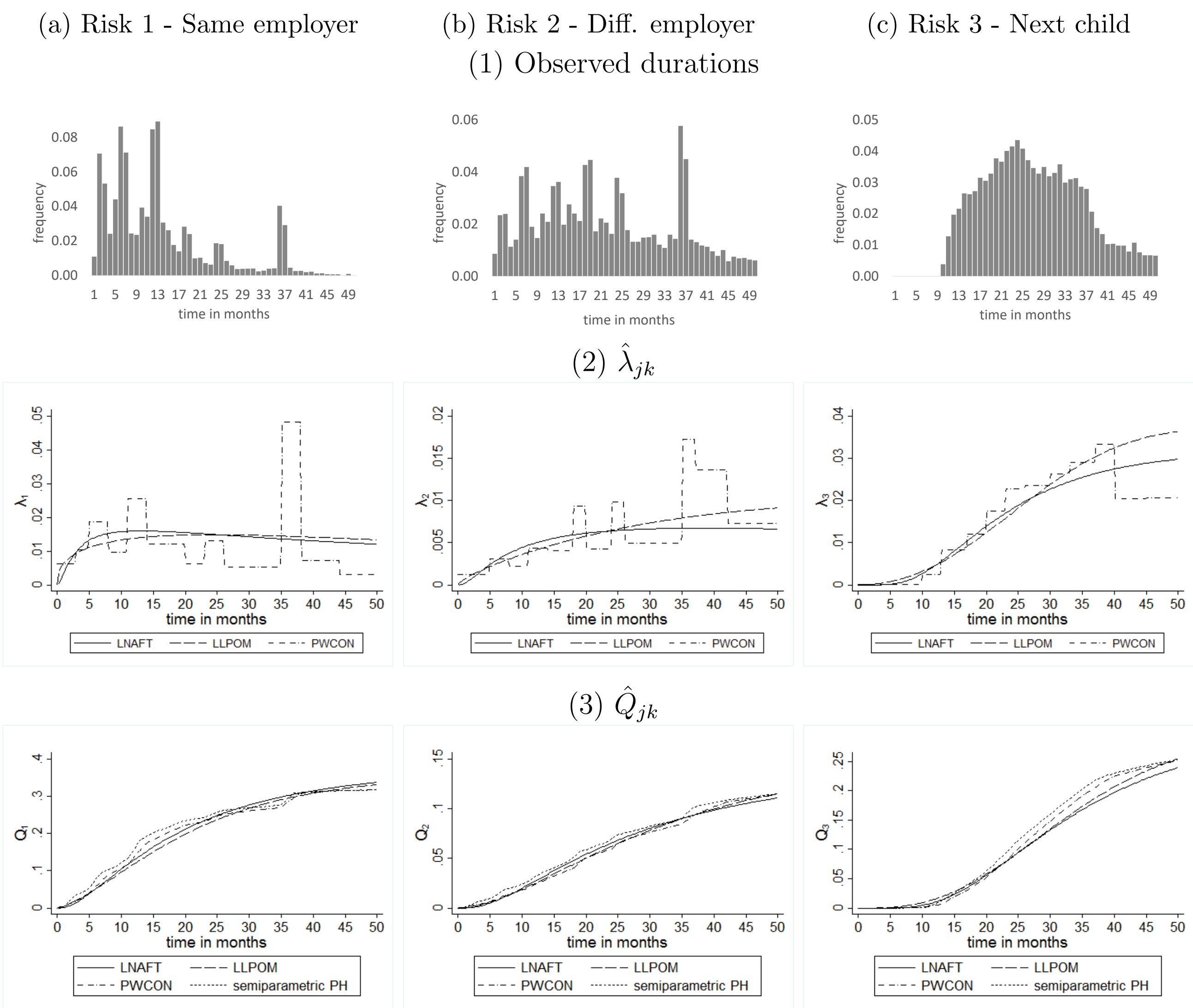

Figure 3: Descriptives and estimation results for risks 1–3.

Notes: Rows (2) and (3): the Gumbel copula, spell-constant parameters, evaluated for the reference mother. Based on German maternity leave data (BASiD).

incidence curves (third row). For the cumulative incidence, the models can be compared to the semiparametric proportional hazard model by Fine and Gray, which does not impose parametric restrictions on the subdistribution baseline hazard functions. Overall, all estimated CICs look similar but there are pronounced differences that stem from the greater flexibility of the piecewise constant and semiparametric model. In particular, for risks 1 and 3, the flexible estimates suggest a 5 percentage points higher CIC than the parametric models for some durations. Overall, the additional flexibility leads to more realistic estimates of the cumulative incidence curves that reflect the existence of mass points.

As suggested in Section 3, we perform pairwise Vuong tests using the piecewise constant cause-specific hazard model to decide which copula captures best the dependence structure between spells. Table 3 reports the sample log likelihood (LL) and test statistic (V) of the pairwise comparisons for the different copulas. For each risk, the Gumbel copula has a higher sample log likelihood than the other two copulas. The Vuong test suggests that the Frank copula and the Clayton copula should be rejected at p-values that are virtually zero. Hence, we maintain the Gumbel copula as preferred copula for all subsequent analyses.

Table 3: Log likelihoods and Vuong test statistics for the Gumbel, Frank and Clayton copulas.

	Risk 1			Risk 2			Risk 3		
	LL	V Gumbel	V Frank	LL	V Gumbel	V Frank	LL	V Gumbel	V Frank
Gumbel	-184,015			-175,721			-176,789		
Frank	-184,340	-13.38		-175,964	-11.04		-177,080	-14.39	
Clayton	-184,321	-8.89	-1.34	-175,944	-6.50	-1.14	-177,060	-8.34	-1.15

Notes: Piecewise constant cause-specific hazard model (PWCON) with spell-constant parameters. Based on German maternity leave data (BASiD).

Next, we use the piecewise constant cumulative incidence curve combined with the Gumbel copula and allow for spell-varying parameters on regressors. In general, any regressor, both constant and spell-varying, can have parameters that differ by spell. We report results for a model with spell-varying parameters on regressors if the parameters were found to significantly vary across spells. For the other variables we reduce the complexity by using a single parameter for all spells. Due to how our sample is constructed, there is no first time mother who has not been employed at time of conception. To avoid multicollinearity, the binary regressor for non-employment is therefore not included for the first spell. Table S8 in the Supplementary Material S.III, columns 1 to 3 show the estimated parameters. Since the model with spell-constant parameters is nested in the model with spell-varying parameters, we conduct likelihood ratio tests to assess whether the additional parameters play a role in the models. For all risks, the likelihood ratio test suggests a significant increase in the sample log likelihood when allowing for spell-varying parameters, see Table 4. Hence, restricting the analysis to the first spell or pooling over multiple spells is not accurate in our application. On average, the Kendall-τs are higher in this specification, ranging from 0.17 to 0.31, indicating misspecification in the models with spell-constant parameters.

To better illustrate the effects of the covariates over time, we compute spell specific partial effects of discrete changes in the regressors on the cumulative incidence curves as described in Section 3. As baseline we use the cumulative incidence of the reference mother, see above. We compare the partial effects of our preferred specification (the Gumbel copula, piecewise constant hazard model with partially spell-varying parameters on regressors) with those obtained for the

Table 4: Likelihood Ratio Test for model with spell-constant parameters vs. spell-varying parameters.

	Risk 1		Risk 2		Risk 3	
	LL	LR stat. *(df)*	LL	LR stat. *(df)*	LL	LR stat. *(df)*
spell-constant only	-184,015		-175,721		-176,789	
spell-varying	-182,641	2748 *(16)*	-175,539	364 *(17)*	-176,300	978 *(17)*

Notes: Piecewise constant cause-specific hazard model (PWCON) with the Gumbel copula. Based on German maternity leave data (BASiD).

semiparametric proportional hazard model (Fine and Gray, 1999) pooled over spells but partially including interaction terms of the regressors and the spell indicator (Table S8 in the Supplementary Material S.III, columns 4-6). Figure 4 shows the partial effects of the employment states ten months prior to birth on CICs. The reference category is full-time employment in the second income tercile. Figure S3 in the Supplementary Material S.III illustrates the partial effects of the other covariates on CICs. Figures S2 and S4 in the Supplementary Material S.III illustrate the according partial effects of the covariates on the hazard function.

The jumps in the cumulative incidence curve for the reference mother observed above (Figure 3) can also be found for the partial effects. For instance, the increase in the propensity to return to the same employer due to having worked full-time with high earnings prior to the third birth jumps from 23 percentage points after 36 months to 27 percentage points after 38 months.

Coefficients vary strongly in size across spells. For all prebirth employment states, the partial effects on the probability of returning to the same employer tend to increase over spells while the partial effects on the probability of changing employer decrease over spells (except for high earnings). Similarly, the partial effects on the propensity to have a next child become smaller with each spell for all prebirth employment states. For instance, the partial effect of receiving a wage belonging to the upper tercile on the propensity to have a next child after 36 months is close to zero after the first birth, -0.10 after the second birth and -0.16 after the third birth.

Even more interestingly, the partial effects also vary in sign across spells. If the mother was part-time employed before the first birth, she is less likely to stay with the same employer. This might be because part-time employment before motherhood is often involuntary. The effect is reversed for the second and third spell: mothers who worked part-time prior to the second and third birth are more likely to return to the same employer, probably because this work arrangement is convenient for child-rearing. Mothers working full-time but earning relatively little show less attachment to the former employer and use the opportunity to change employer after the first childbirth, maybe in the hope of being better paid. After the second and third birth, however, changes in employer become less likely, despite low earnings. The reason might be that changing

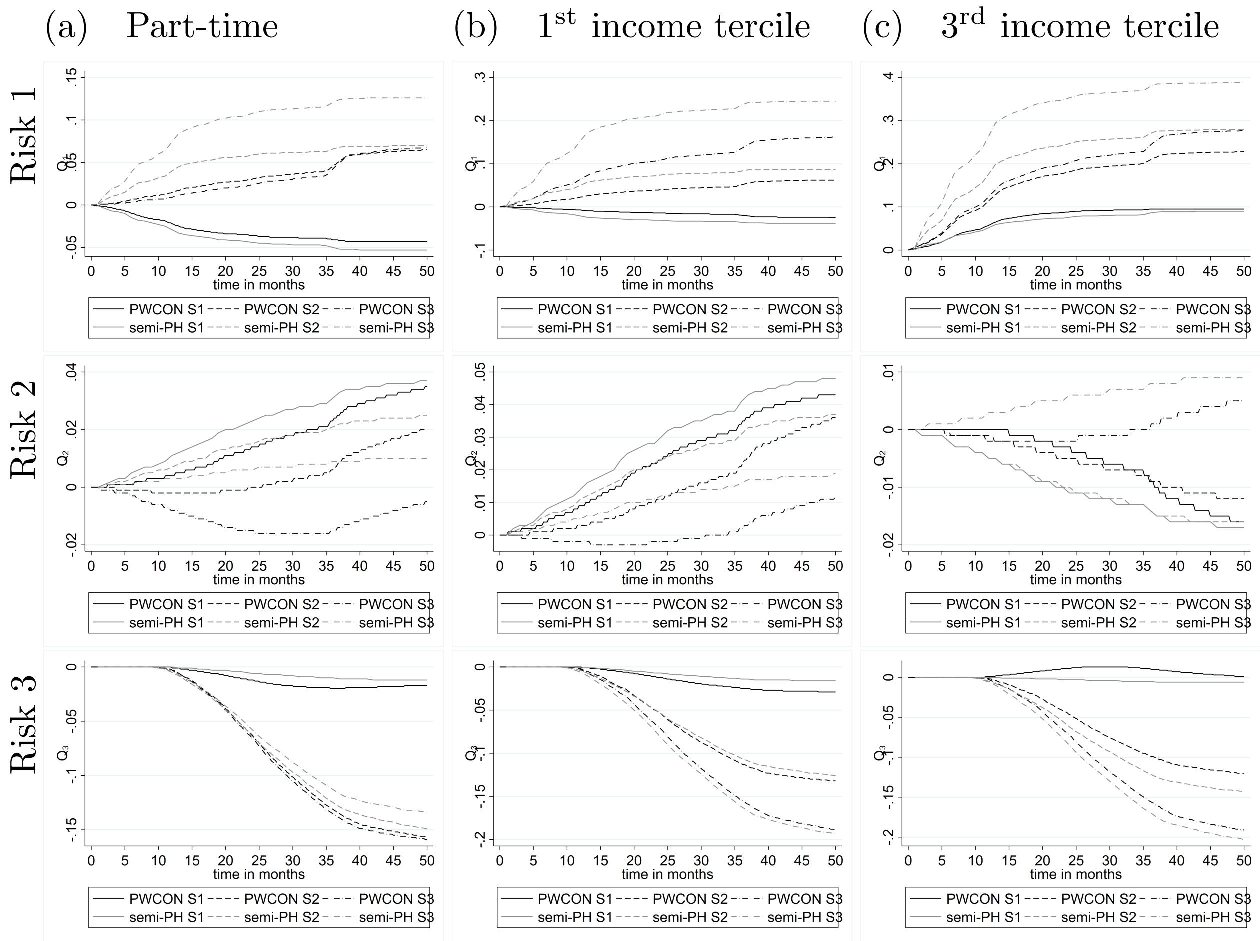

Figure 4: Partial effects of employment states ten months prior to birth on CICs.

Notes: The Gumbel copula, spell-varying parameters. (a) worked part-time; (b) worked full-time with an income in the first income tercile; (c) worked full-time with an income in the third income tercile. S1 - first spell; S2 - second spell; S3 - third spell. Based on German maternity leave data (BASiD).

employer is associated with search and transaction costs which mothers of more than one child cannot or do not want to pay.

The partial effects for the semiparametric PH model and the piecewise constant model partly diverge. The different underlying hazards (subdistribution hazard and cause-specific hazard) lead to different implied CICs and to different restrictions on the partial effects. The difference is often remarkably large. For instance, the difference in the partial effect of earning a high income on the cumulative incidence for risk 1 is 7 percentage points after 36 months for spell 2 and 15 percentage points for spell 3. Even more worrying, the estimated directions of the effects sometimes differ: having no vocational training is estimated to increase the cumulative incidence of risk 1 for spell 3 in the semiparamteric PH model (+3 percentage points after 36 months), while it is the opposite in the piecewise constant model (-3 percentage points after 36 months, see Figure S3 in the Supplementary Material S.III). As pointed out in Section 2, the partial effects for the piecewise constant model can change direction for different t, while the direction is unique for the semiparametric model. Although this is not found to happen often, it is present for example for the effect of part-time work and the third income tercile for risk 2 in the second/third spell. The differences between models are generally the smallest for the first spells, where the number of observations is by far the largest, and greatest for the third spell with fewer observed transitions. Whether the partial effects differ statistically could be investigated by an application of the bootstrap.

Our results should be of interest to employers and policymakers alike. Employers benefit from shorter maternity leaves, e.g. limited losses of human capital during inactivity. Therefore, any variable that increases the cumulative incidence ceteris paribus contributes to a quicker return. Our results show that in particular those with high wages have a much higher likelihood to return to the same employer at a given time than those with a middle wage. The partial effect is in particular pronounced for the second and third spell. The greater probability of returning to a part-time job in the second and third spell also suggests that better compatibility of work and family contributes to return.

Since many years policymakers aim at raising birth rates to counter population aging. Our results confirm the importance of education and wages for having multiple children. Females without educational degree are less likely to have a next child (-2 percentage points after the first birth, -4.5 percentage points after the second birth and -6 percentage points after the third birth after 50 months, see Figure S3 in the Supplementary Material S.III). Low paid mothers are 3 percentage point less likely to have a second child and 13 percentage points less likely to have a third child (after 50 months). Furthermore, governments might also be interested in shortening excessive job protection periods to reduce times of female inactivity and to enlarge labour supply. In the considered German context, this would move the mass points at 36 months to the left and lead to shorter leave periods. Especially quick transitions into an exit state are observed for females with a tertiary education degree and for young women. The partial effect of higher education on risk 1 is large up to month 12 (for risk 2 and 3 up to month 24) and abates thereafter. The same holds for the partial effect of being aged between 18 and 23 years. Especially long maternity leaves are observed for those not employed before birth.

5 Summary and Extensions

This paper studies in detail the nested copula duration model with dependent competing risks and dependent multiple spells. We suggest various practically useful extensions and analyse finite sample performance with Monte Carlo studies. The applicability of the model is shown with an application to large scale maternity leave data. The suggested inference methods can be used to guide the construction of the model, in particular the choice of copula which is typically unknown. It is found that a variety of different effects exists for different risks and spells. The suggested flexible models are able to capture these features and can cope with partly unknown functional forms of the underlying hazards. The piecewise constant cause-specific hazard model outperforms the semiparametric proportional hazard model for the subdistribution in the simulations and allows for richer partial covariate effects that in particular are compatible with changes in the direction of effects for different durations. It should therefore be included in the toolbox of empirical researchers and complement the results of the model by Fine and Gray (1999). We make STATA sample code available to ease its adoption. The suggested model focuses on modelling identifiable quantities such as the cumulative incidences and the multiple spell dependencies. In the case the research interest is on the marginal survivals or the competing risks dependencies, the copula graphic estimator or a direct copula model for the marginal survivals can be applied. While the former requires an assumption about the competing risks copula (Lo et al., 2020), it is weakly identified in the latter approach (Emura et al., 2020).

A disadvantage of the piecewise constant hazard model is that it requires user input about the number of intervals and interval endpoints. Bouaziz and Nuel (2016) suggest an extension to data driven selection of these parameters for the marginal hazard model. A similar extension would be interesting for the cause-specific hazard model.

References

[1] Arntz, M., Dlugosz, S. and Wilke, R.A. (2017) The Sorting of Female Careers after First Birth: A Competing Risks Analysis of Out of Work Duration, *Oxford Bulletin of Economics and Statistics*, 79, 689–716.

[2] Bakoyannis, G. and Touloumi, G. (2012) Practical methods for competing risks data: A review, *Statistical Methods in Medical Research*, 21, 257–272.

[3] Beblo, M., Bender, S. and Wolf, E. (2008) Establishment-level wage effects of entering motherhood, *Oxford Economic Papers*, 61, 11–34.

[4] Beyersmann, J., Latouche, A., Bucholz, A., Schumacher, M. (2009), Simulating competing risks data in survival analysis, *Statistics in Medicine*, 28, 956–971.

[5] Bouaziz, O. and Nuel, G. (2016) L0 regularisation for the estimation of piecewise constant hazard rates in survival analysis, arXiv:1609.04595v2.

[6] Braekers, R. and Veraverbeke, N. (2005) A copula-graphic estimator for the conditional survival function under dependent censoring, *Canadian Journal of Statistics*, 33, 429–447.

[7] Budig, M.J. and England, P. (2001) The wage penalty for motherhood, *American Sociological Review*, 204–225.

[8] Carrière, J.F. (1994) Dependent Decrement Theory. *Transactions of Society of Actuaries*, 46, 45–74.

[9] Colvert, R.E. and Boardman, T.J. (1976) Estimation in the piece-wise constant hazard rate model, *Communications in Statistics - Theory and Methods*, 5, 1013–1029.

[10] Craiu, R.V. and Duchesne, T. (2004) Inference based on the EM algorithm for the competing risks model with masked causes of failure, *Biometrika*, 91, 543–558.

[11] Craiu, R.V. and Lee, T.C.M. (2005) Model Selection for the Competing-Risks Model with and without Masking, *Technometrics*, 47, 457–467.

[12] de Uña-Álvarez, J. and Veraverbeke, N. (2017) Copula-graphic estimation with left-truncated and right-censored data, *Statistics*, 51, 387–403.

[13] Emura, T., Matsui, S. and Rondeau, V. (2019) *Survival Analysis with Correlated Endpoints: Joint Frailty-Copula Models*, JSS Research Series in Statistics, Springer.

[14] Emura, T., Shih, J.H., Ha, I.D. and Wilke, R.A. (2020) Comparison of the marginal hazard model and the subdistribution hazard model for competing risks under an assumed copula. *Statistical Methods in Medical Research*, 29, 2307–2327.

[15] Fine, J. P. and Gray, R. J. (1999) A proportional hazard model for the subdistribution of a competing risk, *Journal of the American Statistical Association*, 94, 496–509.

[16] Fitzenberger, B., Steffes, S. and Strittmatter, A. (2016) Return-to-job during and after parental leave *The International Journal of Human Resource Management*, 27(8), 803–831.

[17] Gangl, M., and Ziefle, A. (2009) Motherhood, labor force behavior, and women?s careers: An empirical assessment of the wage penalty for motherhood in Britain, Germany, and the United States, *Demography*, 46(2), 341–369.

[18] Ha, I.D., Kim, J.M. and Emura, T. (2019) Profile likelihood approaches for semiparametric copula and frailty models for clustered survival data. *Journal of Applied Statistics*, 46, 2553–2571.

[19] Hochfellner, D., Müller, D. and Wurdack, A. (2012) Biographical data of social insurance agencies in Germany: Improving the content of administrative data. *Journal of Contextual Economics: Schmollers Jahrbuch*, 132(3), 443–451.

[20] Joe, H. (2005) Asymptotic efficiency of the two-stage estimation method for copula-based models, *Journal of Multivariate Analysis*, 94, 401–419.

[21] Kyyrä, T. (2009) Marginal Effects for Competing Risks Models with Piecewise Constant Hazards, *Oxford Bulletin of Economics and Statistics*, 71, 539–565.

[22] Lancaster, T. (1990) *The Econometric Analysis of Transition Data*, Econometric Society Monographs, 17, Cambridge, UK.

[23] Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data*, 2nd ed. New York: Wiley.

[24] Lo, S.M.S. and Wilke, R.A. (2010) A copula model for dependent competing risks, *Journal of the Royal Statistical Society Series C*, 59, 359–376.

[25] Lo, S.M.S. and Wilke, R.A. (2014) A regression model for the Copula-Graphic Estimator. *Journal of Econometric Methods*, 3, 20–41.

[26] Lo, S.M.S., Mammen, E. and Wilke, R.A. (2020) A nested copula duration model for competing risks with multiple spells, *Computational Statistics and Data Analysis*, 150.

[27] Nelsen, R.B. (2006), *An Introduction to Copulas*, 2nd Edition, Springer, New York.

[28] Rivest, L. and Wells, M.T. (2001) A Martingale Approach to the Copula-Graphic Estimator for the Survival Function under Dependent Censoring, *Journal of Multivariate Analysis*, 79, 138–155.

[29] Rodrigues, B. and Vergnat,V. (2019) The time and the transitions back to work in France after maternity, *Review of Economics of the Household*, 17(3), 861–888.

[30] Schönberg, U. and Ludsteck, J. (2007) Maternity leave legislation, female labor supply, and the family wage gap. IZA Discussion Paper No. 2699.

[31] Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges", Publ. Inst. Statist. Univ. Paris, 8, 229–231.

[32] Trivedi, P.K. and Zimmer, D.M. (2005) Copula Modeling: An Introduction for Practicioners, *Foundations and Trends in Econometrics*, 1, 1–111.

[33] Vuong, Q. (1989) Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57, 307–33.

[34] Waldfogel, J. (1997) The effect of children on women's wages. *American Sociological Review*, 209–217.

[35] Zheng, M. and Klein, J.P. (1995), Estimates of marginal survival for dependent competing risks based on assumed copula. *Biometrika*, 82, 127–138.

Appendix

Proof of Proposition 2

To simplify the notation, we omit index k and $\boldsymbol{\beta}_k$. For the same reason, we treat all relevant functionals to be sufficiently smooth such that they are differentiable and integrable in t and differentiable in $\boldsymbol{x}$. For example the PWCON model has a discontinuous cause-specific hazard in t and would require a combination of summation and integration (see Kyrrä, 2009). Without that this changes the nature of the results. Let

$$f_j(t;\boldsymbol{x}) \;=\; \lambda_j(s;\boldsymbol{x})S(s;\boldsymbol{x})$$

and therefore $Q_j(t;\boldsymbol{x}) = \int_0^t f_j(s;\boldsymbol{x})ds$. We assume $f'_{jl}(t;\boldsymbol{x}) = \partial f_j(t;\boldsymbol{x})/\partial x_l$ is bounded and there exists an interval for t such that $f'_{jl}(t;\boldsymbol{x})$ has the same sign for all $t \in [t_1, t_2)$.

We have

$$\begin{aligned} Q'_{jl}(t_2;\boldsymbol{x}) &= \int_0^{t_2} f'_{jl}(s;\boldsymbol{x})\ ds \\ &= \int_0^{t_1} f'_{jl}(s;\boldsymbol{x})\ ds + \int_{t_1}^{t_2} f'_{jl}(s;\boldsymbol{x})\ ds. \end{aligned}$$

The sign of $Q'_{jl}(t_1;\boldsymbol{x})$ is different from $Q'_{jl}(t_2;\boldsymbol{x})$ if

$$\begin{aligned} &(C1) \qquad \left|\int_0^{t_1} f'_{jl}(s;\boldsymbol{x})\ ds\right| < \left|\int_{t_1}^{t_2} f'_{jl}(s;\boldsymbol{x})\ ds\right|, \text{and} \\ &(C2) \qquad sign\left(\int_0^{t_1} f'_{jl}(s;\boldsymbol{x})\ ds\right) \neq sign\left(\int_{t_1}^{t_2} f'_{jl}(s;\boldsymbol{x})\ ds\right). \end{aligned}$$

Without loss of generality, we assume that

$$\int_0^{t_1} f'_{jl}(s;\boldsymbol{x})\ ds > 0$$

and

$$\int_{t_1}^{t_2} f'_{jl}(s;\boldsymbol{x})\ ds < 0.$$

The above conditions $C1$ and $C2$ can be combined as

$$(C3) \qquad \int_0^{t_1} f'_{jl}(s;\boldsymbol{x})\ ds < -\int_{t_1}^{t_2} f'_{jl}(s;\boldsymbol{x})\ ds.$$

Let $s_1 = \arg\sup_{t\in[0,t_1)} f'_{jl}(s;\boldsymbol{x})$ and $s_2 = \arg\sup_{t\in[t_1,t_2)} f'_{jl}(s;\boldsymbol{x})$. Then, $f'_{jl}(s_2;\boldsymbol{x}) < 0$ and takes the least negative value of $f'_{jl}(t;\boldsymbol{x})$ for $t \in [t_1, t_2)$. We therefore have

$$\begin{aligned} 0 < \int_0^{t_1} f'_{jl}(s;\boldsymbol{x})\ ds &\leq \int_0^{t_1} f'_{jl}(s_1;\boldsymbol{x})\ ds \equiv t_1 f'_{jl}(s_1;\boldsymbol{x}), \text{ and} \\ 0 < -(t_2 - t_1) f'_{jl}(s_2;\boldsymbol{x}) &\equiv -\int_{t_1}^{t_2} f'_{jl}(s_2;\boldsymbol{x})\ ds \leq -\int_{t_1}^{t_2} f'_{jl}(s;\boldsymbol{x})\ ds. \end{aligned}$$

Hence, C3 holds if there existed t_1 and $\epsilon = t_2 - t_1 > 0$ such that :

$$
\begin{aligned}
0 &< t_1 f'_{jl}(s_1; \boldsymbol{x}) < -\epsilon f'_{jl}(s_2; \boldsymbol{x}), \\
0 &< \frac{t_1}{\epsilon} f'_{jl}(s_1; \boldsymbol{x}) < -f'_{jl}(s_2; \boldsymbol{x}).
\end{aligned} \tag{21}
$$

(21) holds if (i) $f'_{jl}(s_2; \boldsymbol{x}) < 0$ and (ii) $|f'_{jl}(s_2; \boldsymbol{x})| > c f'_{jl}(s_1; \boldsymbol{x}) > 0$ for some $c > 0$. c is small if t_1 is small or $t_2 - t_1$ is large. It depends on the model and covariate x_l, whether these conditions are satisfied or not. This is illustrated and discussed in more detail in what follows.

Note that

$$
\begin{aligned}
f'_{jl}(s; \boldsymbol{x}) &= \lambda_j(s; \boldsymbol{x}) S'(s; \boldsymbol{x}) + \lambda'_{jl}(s; \boldsymbol{x}) S(s; \boldsymbol{x}) \\
&= -f_j(s; \boldsymbol{x})(\Lambda'_{1l}(s; \boldsymbol{x}) + \Lambda'_{2l}(s; \boldsymbol{x})) + \lambda'_{jl}(s; \boldsymbol{x}) f_j(s; \boldsymbol{x}) / \lambda_j(s; \boldsymbol{x}) \\
&= -f_j(s; \boldsymbol{x})[\Lambda'_{1l}(s; \boldsymbol{x}) + \Lambda'_{2l}(s; \boldsymbol{x}) - \lambda'_{jl}(s; \boldsymbol{x}) / \lambda_j(s; \boldsymbol{x})].
\end{aligned}
$$

Condition (i) $f'_{jl}(s_2; \boldsymbol{x}) < 0$ requires

$$
\Lambda'_{1l}(s_2; \boldsymbol{x}) + \Lambda'_{2l}(s_2; \boldsymbol{x}) > \Lambda'_{jl}(s_2; \boldsymbol{x}) / \lambda_j(s_2; \boldsymbol{x}), \tag{22}
$$

which may be the case or not. Take as an example a proportional hazard model $\Lambda_j(t; \boldsymbol{x}) = \Lambda_{j0}(t)\phi_j(\boldsymbol{x})$, $\lambda_j(t; \boldsymbol{x}) = \lambda_{j0}(t)\phi_j(\boldsymbol{x})$ and $\phi_j(\boldsymbol{x}) > 0$. In this case (22) becomes

$$
\Lambda_{10}(s_2)\phi'_{1l}(\boldsymbol{x}) + \Lambda_{20}(s_2)\phi'_{2l}(\boldsymbol{x}) \quad > \quad \phi'_{jl}(\boldsymbol{x}) / \phi_j(\boldsymbol{x}).
$$

This likely holds for s_2 large as $\Lambda_{j0}(s)$ grows to ∞ as t increases, whenever $\phi'_{jl}(\boldsymbol{x}) > 0$ for $j = 1, 2$. (i) therefore cannot be ruled out even under strong restrictions on the role of the covariates. For condition (ii) to hold, we need for some $c > 0$

$$
\begin{aligned}
& |f_j(s_2; \boldsymbol{x})[\Lambda'_{1l}(s_2; \boldsymbol{x}) + \Lambda'_{2l}(s_2; \boldsymbol{x}) - \Lambda'_{jl}(s_2; \boldsymbol{x}) / \lambda_j(s_2; \boldsymbol{x})]| \\
> \; & c|f_j(s_1; \boldsymbol{x})[\Lambda'_{1l}(s_1; \boldsymbol{x}) + \Lambda'_{2l}(s_1; \boldsymbol{x}) - \lambda'_{jl}(s_1; \boldsymbol{x}) / \lambda_j(s_1; \boldsymbol{x})]|.
\end{aligned}
$$

Again, this may be true or not. Take again as an example $\Lambda_j(t; \boldsymbol{x}) = \Lambda_{j0}(t)\phi_j(\boldsymbol{x})$, the above condition becomes

$$
\begin{aligned}
& |f_j(s_2; \boldsymbol{x})[\Lambda_{10}(s_2)\phi'_{1l}(\boldsymbol{x}) + \Lambda_{20}(s_2)\phi'_{2l}(\boldsymbol{x}) - \phi'_{jl}(\boldsymbol{x}) / \phi_j(\boldsymbol{x})]| \\
> \; & c|f_j(s_1; \boldsymbol{x})[\Lambda_{10}(s_1)\phi'_{1l}(\boldsymbol{x}) + \Lambda_{20}(s_1)\phi'_{2l}(\boldsymbol{x}) - \phi'_{jl}(\boldsymbol{x}) / \phi_j(\boldsymbol{x})]|.
\end{aligned}
$$

This generally holds if $f_j(s_2; \boldsymbol{x}) > c f_j(s_1; \boldsymbol{x})$ since by definition $\Lambda_{j0}(s_2) > \Lambda_{j0}(s_1)$ for all $s_2 > s_1$. Condition (ii) therefore holds for c small enough. We have therefore illustrated that even for the proportional hazards model, where the direction of the partial effect on the hazard is unique for all t, there is no analogous result for $Q'_{jl}(t; \boldsymbol{x})$ as the direction of partial covariate effect is not restricted.